

The Imaging of  
Extra-Galactic Low-Energy Gamma-Ray Sources  
Prospects, Techniques, and Instrumentation

Thesis by  
Mark Harold Finger

In Partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Philosophy

California Institute of Technology  
Pasadena, California

Submitted August 18, 1987

SRL 87-21

## Acknowledgements

I would like to extend thanks to the following people who supported me, guided me and helped me through my dissertation. First I must thank my advisor Thomas Prince for his faith in me, his assistance and his great enthusiasm. I would also like to extend thanks to Tom's wife Charlene Reichert for her warmth and her fine dinner parties. I would like to thank Ed Stone for his support and inspiration. Fellow lab-workers and technicians who were there when I needed them were Rick Cook, Bill Althouse, Allen Cummings, Steve Schindler, Chris Starr and Jim Wegger. Special thanks go to Deeby Kadrie for excellent work and secretarial support.

My parents, Paul and Shirley Finger, have been encouraging throughout, intellectually, emotionally and financially. My special friends Mark and Alice Cronin-Golomb, Debbie Vigil and Geoff Blewitt deserve my thanks for their concern, encouragement and good-natured patience.

Finally, I must thank my wife Lynn for her love, patience and support, which have made this work possible. I thank her for her faith in me.

This work was supported in part by NASA grant NGR 05-002-160.

### Abstract

The handful of extra-galactic low-energy gamma-ray sources so far observed are all active galaxies, which are expected to dominate future detections. The nature of these compact, highly luminous sources is at present not clear; however, they may be powered by massive black holes. Many of these sources may produce their peak luminosity in the 0.5 to 5.0 MeV energy band, and observation in this energy range will be important in revealing the nature of their central power-house.

Improved understanding of the nature of active galaxies will require detailed observations of 10-20 sources, while understanding of their gamma-ray luminosity function and its evolution will require the detection of  $\sim 100$  sources. From x-ray number counts and the presently available information about active galaxy spectra, we estimate the hard x-ray and low-energy gamma-ray number source-flux relation  $N(>S)$  for active galaxies. Instruments capable of detecting  $\sim 100$  active galaxies at low-energy gamma-ray energies are achievable. These instruments will, however, be observing sources with fluxes some  $10^{-3}$  -  $10^{-4}$  times lower than their instrumental background level, and will require careful control of systematic errors.

The angular resolution of an instrument, as well as its sensitivity, can limit the number of sources it can observe. We present an investigation of the angular resolution requirements for future low-energy gamma-ray

instruments. We find that the strictest requirements arise not from the need to resolve detectable sources, but from the need to control the level of direction-to-direction fluctuations in the diffuse background level. We conclude that gamma-ray instruments capable of detecting 100 active galaxies must have sub-degree angular resolution.

We propose use of the coded aperture imaging technique as a method of achieving accurate control of systematic errors and fine angular resolution without unduly increasing the time needed to conduct full-sky surveys. This is a technique that employs a partially opaque mask to spatially modulate the source flux incident upon a position-sensitive photon detector. We present an analysis of coded aperture imaging for instruments that employ masks based on hexagonal uniformly redundant arrays. Rotation of such a mask allows complete, position-by-position background subtraction on short time-scales, and removes the periodic ambiguity inherent in uniformly redundant arrays.

An instrument, the Gamma-ray Imaging Payload, has been built that employs these imaging techniques. The primary detector of the instrument is a 41cm diameter by 5 cm thick NaI(Tl) Anger camera. We describe the design and testing of the instrument in detail. Preliminary results from a balloon flight of the instrument are shown, demonstrating its imaging performance.

## Table of Contents

Section	Title	Page
1	Extra-Galactic Gamma-Ray Astronomy	1
1.1	Current Observational Status	4
1.1.1	Active Galactic Nuclei	5
1.1.2	The X-ray and Gamma-ray Spectrum of AGN	8
1.1.3	The Diffuse X-ray and Gamma-ray Background	13
1.2	Source Models	23
1.3	Future Expectations	32
1.3.1	Estimates of $N(>F)$	33
1.3.2	Instrumental Requirements	42
2	Angular Resolution Requirements for Gamma-ray Instruments	44
2.1	Source Localization	44
2.2	Source Confusion	46
2.2.1	Source Confusion and Background Fluctuation Estimates	47
2.2.2	The Deflection Distribution	50
2.2.3	Source Detection	64
2.3	Conclusions	80
3	Gamma-Ray Imaging	82
3.1	Coded Aperture Imaging	85
3.2	The Instrument Model	91
3.3	Source Flux and Direction Estimation	94
3.4	Source Flux and Location Errors	101

3.5	Continuous Mask Rotation	107
3.6	Source Detection and Noise Peaks	111
4	The Gamma-Ray Imaging Payload (GRIP)	127
4.1	Overview	127
4.2	The Imaging System	134
4.2.1	The Mask	135
4.2.2	The Primary NaI(Tl) Detector	140
4.2.3	The Calibration Systems	177
4.2.4	Possible Design Improvements	179
4.2.5	The Shield System	179
4.3	Support Sub-Systems	184
4.3.1	Event Logic	190
4.3.2	The Rate System	196
4.3.3	The Telemetry and Recording Systems	197
4.3.4	The Pointing System	198
4.3.5	Thermal Control Systems	200
5	Epilogue	204
A.	Appendix: Properties of the Image Function	216
	References	220

## List of Figures

Figure	Title	Page
1.1	The Diffuse Cosmic Background Spectrum	15
1.2	X-ray Luminosities and Minimum Variability	26
	Time-Scales for AGN	
1.3	AGN $N(>S)$ at 100 keV	35
1.4	AGN $N(>S)$ at 1 MeV	37
2.1	Fluctuation Distributions for Various Values of $\gamma$	55
2.2	Fluctuation Distributions for Various Values of $R$	57
2.3	Model Response Function	63
2.4	Required Collimation versus Sensitivity	66
2.5	The Maximum Intensity Distribution for a Large Deflection	71
2.6	Maximum Intensity Distributions for Small Deflections	73
2.7	Distribution of the Ratio of Maximum Source Intensity to Deflection for Small Deflections	76
2.8	Survey Completeness versus Contamination	79
3.1	The HURA of order 127	88
3.2	Imaging Instrument Geometry	93
3.3	Source Flux and Locations Errors versus Cell Size	106
3.4	Extension of the Field of View Through Mask Rotation	109
3.5	Noise Peak Distributions	119
3.6	Integral Noise Peak Distributions	121
3.7	The Effect of Image Searching on Source Detection	125

4.1	Basic Elements of the GRIP Instrument	129
4.2	GRIP Flight Configuration	133
4.3	The GRIP Mask Pattern	137
4.4	Opacity of Lead	139
4.5	NaI Photo-Peak Efficiency	143
4.6	Primary Detector Assembly	151
4.7	Detector Response Measurement Grid	156
4.8	The Primary Detector Central PMT Response	160
4.9	Calibration Hole Plate Data	165
4.10	The Detector's Point Spread Function at 122 keV	168
4.11	The Detector's Point Spread Function at 662 keV	170
4.12	The Detector's Spatial Resolution as a Function of Energy	172
4.13	Depth Discrimination	176
4.14	The Effect of the Shield Veto on the Energy Loss Spectrum	186
4.15	GRIP Organization	189
4.16	Event Logic Flow Chart	193
5.1	GRIP In Flight Background Spectrum	207
5.2	GRIP Thorium Spectrum	210
5.3	GRIP Crab Region Image	212
5.4	GRIP Cygnus Region Image	214



## 1. Extra-Galactic Gamma-Ray Astronomy

### Introduction

Gamma-ray astronomy is still in its infancy, with extra-galactic observation only begun. At present, the diffuse background and only a handful of extra-galactic objects have been detected above 0.5 MeV. However, the picture that emerges from these observations and the inferences from other energy bands suggest that important contributions to astronomy will emerge from future gamma-ray observations.

The objects detected so far are nearby representatives of various classes of active galaxies. The active galaxies, including Seyfert galaxies, QSOs, radio galaxies, and BL Lacertae objects, have in common a compact core or nucleus undergoing a violent release of energy. Such active galactic nuclei (AGN) have been studied extensively from radio to x-ray energies. However it is now clear that for many AGN the peak luminosity occurs in the low-energy gamma-ray range near 1 MeV. Low energy gamma-rays ( $\sim 0.5$  to 10 MeV) are the most penetrating form of electro-magnetic radiation. Thus many AGN release a major portion of their power in the radiation most revealing of their central power source. Observations at these energies will therefore be essential for understanding the nature of these objects.

AGN have been found to be variable in the optical, x-ray and gamma-ray ranges. The time-scale for this variability is limited by the size of the

emitting region, with small regions capable of more rapid variation. Thus study of the variability of AGN at gamma-ray energies will be revealing of the size of the central source, providing important constraints for AGN models.

As the sensitivities of instruments improve and systematic surveys are conducted, the number of detected extra-galactic sources should increase dramatically. With a few tens of sources, statistical studies of the distribution of AGN in luminosity and redshift can be used to look for evidence of the evolution of the source population. Such studies have been conducted at other energies and have revealed that active galaxies were on average more luminous in the past (e.g. Schmidt and Green 1986, Maccacaro et al. 1984). However the distribution in luminosity and the degree to which it has evolved is different in each energy band. As yet there is no clear understanding of how the luminosity distributions at different energies are linked, or what the evolution implies about the history of individual sources. Study of the distribution of gamma-ray luminosities, which are closely linked to the total power output of the sources, should help clarify this situation.

At present there is no clear understanding of the origin of the diffuse cosmic background at high energies. The measured isotropy at x-ray energies, as well as arguments pertaining to energetics, lead to the conclusion that it is truly of extra-galactic origin (Gould 1967). As such it may contain information about the universe in very remote epochs. Whether it is due to the superposition of unresolved discrete sources or to a diffuse mechanism operating in the intergalactic medium, or both, has been debated since its initial detection. Historically both theories have come in and out of vogue. However

present estimates of the contribution due to AGN in the low-energy gamma-ray range indicate that AGN may account for the entire flux. In fact, there is difficulty in limiting the estimated summed AGN flux to below the level of the background. In the range from 0.5 to 5 MeV the cosmic background spectra flattens, and this feature is often taken as evidence that this portion of the background is due to AGN (Rothschild et al. 1983). However such estimates rely on extrapolation from x-ray observations, and it is possible that the spectral extrapolation used is valid for only a small fraction of the sources. Clearly only observations of AGN in the low-energy gamma-ray range will settle the issue of their contribution to the background in this region of the spectrum.

In Sections 1 and 2 of this chapter we present some the results of past extra-galactic observations in low-energy gamma-rays, and review the scientific issues involved with these observations. We then turn in to the expected results of future observations in Section 3. Our main interest here is to outline the course of future observations, keeping the requirements on future gamma-ray instruments in mind. To this end we present an estimate of the number of extra-galactic gamma-ray sources which will be detected at a given instrumental sensitivity, and begin a discussion of the angular resolution requirements of useful future gamma-ray survey instruments.

Spectral observations of x-ray sources are often presented in terms of the incident energy flux, while gamma-ray observations are more often presented in terms of the incident number flux. To avoid confusion, we will use the symbol  $S$  to indicate an energy flux ( $\text{erg/cm}^2\cdot\text{s}\cdot\text{keV}$ , or  $\text{erg/cm}^2\cdot\text{s}$  over a given energy band) while using the symbol  $F$  to indicate a number flux

(photons/cm<sup>2</sup>·s·keV). For power law spectra we will refer to the energy spectral index  $\alpha$  with  $S(E) \propto E^{-\alpha}$ , and the number spectral index  $\Gamma = \alpha + 1$ , with  $F(E) \propto E^{-\Gamma}$ .

### 1.1 Current Observational Status

At present only a small number of extra-galactic sources have been detected in low-energy gamma-rays. The extragalactic sources so far detected in the 500 keV to 10 MeV energy range are the Seyfert galaxies NGC4151 and MCG8-11-11, and the radio galaxy Centaurus A (NGC5128). In addition the MISO low-energy gamma-ray telescope has observed a source in the region of the COS-B source CG135+1 which may be either QSO 0241+622 or the radio galaxy GT0236+61 (Perotti et al. 1980).

These sources are observed to have rather flat spectra, with a photon number power law index of  $\Gamma \approx 1-1.8$  above 100 keV. The spectrum of MGC8-11-11 is observed to break to a power law with index  $\Gamma \approx 3$  at approximately 3 MeV.

The detection of these sources could have been expected. These sources can be classified AGN, which have been extensively studied at x-ray energies. The extra-galactic sources observed in the x-ray fall mainly into two groups, galactic clusters and AGN. The x-rays from galactic clusters have a thermal bremsstrahlung spectra with a typical temperature of  $\sim 10$  keV. Clusters are therefore not expected to be observable at higher energies. The AGN consist of Seyfert galaxies, QSOs, radio galaxies, and BL Lac objects, and are observed to have a power law spectra in the x-ray region.

### 1.1.1 Active Galactic Nuclei

Active galaxies are characterized by luminous, star-like nuclei, whose optical spectra consist of a featureless non-thermal continuum underlying emission lines from a variety of ionization states.

**Seyfert Galaxies.** Many of the active galaxies detected at x-ray energies are Seyfert galaxies, which are characterized by an optically compact nucleus which emits intense broad emission lines. Seyferts are classified as type I or type II on the basis of the widths of the allowed and forbidden emission lines. Type I Seyferts have broad Balmer lines, and narrow, strong forbidden lines. Type II Seyferts have narrow Balmer lines and narrow, weak forbidden lines. Type I Seyferts have been found to be stronger x-ray sources. The broad wings on the permitted lines in type I Seyferts indicate velocities in the emitting gas on the order of  $10^4$  km/s. The lack of broad wings in the forbidden lines indicates that the forbidden lines come from a separate region and that the gas density in the Broad Line Region (BLR) is dense enough ( $\sim 10^8/\text{cm}^3$ ) that the emission of broadened forbidden lines is suppressed. Observation of forbidden radiation in nearby Seyferts has revealed distinct clouds with radii of a few parsecs, while variability in the Balmer lines indicates that the BLR is smaller than 0.1 parsec (Ulrich 1985).

The gamma-ray and x-ray emission of Seyfert I galaxies are also observed to be variable. The low-energy gamma-ray flux from NGC4151 has been observed to vary by as much as a factor of ten, with a variability time scale of  $\sim 6$  months (Bassani and Dean 1983). This sets an upper limit of 0.15pc to the size of the low-energy gamma-ray emitting region. Flares in the 2-10 keV emission of NGC4151 have been seen with rise times of a half day

or less (Lawrence 1980), indicating that the x-ray source is  $10^{15}$  cm or less in radius. The x-ray variability of Seyferts is found to correlate with their luminosity; on a time-scale of six months, more low-luminosity sources are found to be variable than high-luminosity sources, with the low luminosity sources showing the highest degree of variability (Mushotzky 1982). Time-scales between six months and a week are not well sampled but for time-scales less than a week only a few low-luminosity Seyfert I's have been detected to be variable (Tennant and Mushotzky 1983). Most notable of these is NGC 6814 which has shown a ten-fold variation in intensity in a few minutes with no apparent change in spectral index.

**QSOs.** The defining characteristics of a Quasi-Stellar Object (QSO) were for many years taken to be a stellar appearance, a strong ultra-violet excess, variable optical emission, an emission line spectra with broad permitted lines, and large redshifts (Hazard & Mitton 1977). Recent work has, however, thrown this definition into disarray. Much of this is due to the detection of numerous nearby low-luminosity quasars by the Einstein x-ray observatory.

Seyfert galaxies have the emission line spectra typical of a QSO, so that a low-redshift object considered to be a Seyfert galaxy would at higher redshifts be considered a QSO. A study of quasar morphology (Malkan et al. 1984) found that 15 of 24 x-ray selected quasars in fact had extended emission consistent with the starlight of a surrounding galaxy. The multi-color photometry of x-ray selected quasars reveals a continuous distribution in B-V, indicating a population of QSOs with no UV excess (Gioia et al. 1984). Many authors are therefore of the opinion that the distinction between QSOs and

Seyfert Type I is merely a subdivision into two luminosity classes (e.g. Setti 1984, Maccacaro 1984).

**Radio Galaxies.** Radio Galaxies are characterized by radio emitting structures, which extend from tens to thousands of kiloparsecs from a central elliptical galaxy. The radio emission, which is generally polarized and of a power law form ( $\alpha \approx 0.6$ ) is probably synchrotron radiation from relativistic electrons accelerated *in situ*.

Radio galaxies are often x-ray sources. An x-ray survey of a sample of radio galaxies from the 3CR catalog found 26 out of 43 galaxies to have detectable emission in the 0.5 to 3 keV energy band (Fabbiano et al. 1984). These galaxies tended to be at the center of high-luminosity double-lobe radio sources with detectable nuclear radio emission, and showed emission lines in their optical spectra similar to Seyfert galaxies. The x-ray emission for these sources was consistent with point sources, and the x-ray luminosities were highly correlated with the 5 GHz nuclear radio luminosities.

One of the galaxies detected in low-energy gamma-rays is the radio galaxy CEN A. This object was observed by Hall et al. (1976) in the 30 keV to 12 MeV energy range. The continuum spectra was found to be well fit by a power law with spectral index  $\Gamma = 1.90 \pm 0.04$ . They also reported the detection of several line features. CEN A has been more recently observed by the A4 instrument on HEAO-1 (Baity et al. 1981). The continuum in the 80 keV to 2.3 MeV range was consistent with a power law with spectral index  $\Gamma = 1.65 \pm 0.04$ , while the previously reported line features were not detected.

**BL Lacertae Objects.** BL Lacs have non-thermal optical continuum spectra steeper than that of QSOs, lack the emission lines characteristic of

QSOs and Seyfert galaxies, and are characterized by rapidly variable optical flux with strong and variable polarization (Hazard and Mitton, 1977). They are relatively rare in both the optical and x-ray band.

There are only five objects with well-studied x-ray spectra (Urry 1984), and there is little information about their hard x-ray spectra. In x-rays they are highly variable with a 20-fold increase in the 2-10 keV flux of Mkn 421 reported over a period of ten days (Ricketts et al. 1976). The spectral indices of BL Lacs in the 2-10 keV energy band range from 1.3 to 5 with spectral index variations being detected in some sources (Maccacaro and Gioia, 1983). The 2-50 keV spectrum of Mkn 421 and Mkn 501 are consistent with a power law with a number spectral index of  $\Gamma = 1.7$  (Dean and Ramsden 1981).

### 1.1.2 The X-ray and Gamma-ray Spectrum of AGN

Active galaxies are observed to have rather hard x-ray spectra and thus are expected to be observable at gamma-ray energies. X-ray observations have lead to the conclusion that AGN have power law hard x-ray number spectrum with spectral index  $\Gamma$  compatible with a universal value of 1.7 (see below). This consistency over a wide range of luminosities points to a common, fairly scale invariant, emission process. For such a spectrum the luminosity per decade in energy increases with energy, so that we expect some break or turnover in the spectrum at higher energies. Upper limits from SAS-2 and COS-B for AGN (Bignami et al. 1979, Pollock et al. 1981) lead to the conclusion that in general this break occurs below 50 MeV. Thus the high-energy luminosity of these objects generally peaks somewhere between 100 keV and 50 MeV.



The largest body of published AGN x-ray spectra come from the satellites HEAO-1 and HEAO-2 (the Einstein Observatory). Table 1.1 summarizes the characteristics of the A-2 and A-4 experiments on board HEAO-1, and the Imaging Proportional Counter (IPC) and Solid State Spectrometer (SSS) on board HEAO-2.

Satellite	HEAO 1		Einstein Observatory	
Operational	1977-1979		1978-1981	
Experiment	A-2	A-4	SSS	IPC
Type	Gas Proportional Counters	NaI/CsI Phoswich Scintillators	Silicon(Li) Solid State Detector	Gas Proportional Counter
Energy Range	2-60 keV	12-2300 keV	0.4-4 keV	0.15-4 keV
Field of View	$1.5^{\circ} \times 3^{\circ} - 3^{\circ} \times 6^{\circ}$	$1.5^{\circ} \times 20^{\circ}$	6' Diameter	$75' \times 75'$
Beam Size	$1.5^{\circ} \times 3^{\circ} - 3^{\circ} \times 6^{\circ}$	$1.5^{\circ} \times 20^{\circ}$	6' Diameter	$1' \times 1'$
Reference	Rothschild et al. (1979)	Matteson et al. (1978)	Giacconi et al. (1979a)	Giacconi et al. (1979a)

**Table 1.1.** Characteristics of selected experiments from the High Energy Astronomy Observatory satellites HEAO 1 and HEAO 2 (the Einstein Observatory).

The instruments of HEAO-1 had coarse angular resolution achieved with passive collimation and together covered a rather broad energy range. In contrast the Einstein Observatory was equipped with grazing incidence angle mirrors providing fine angular resolution within a rather restricted energy range. The focusing optics of the Einstein Observatory allowed an improvement in signal-to-noise ratio by providing a large collection area to rather small detectors. This was best utilized by the position-sensitive IPC, which detected numerous sources unobserved by HEAO-1 A2. The SSS which had superior energy resolution, lacked position resolution, and therefore could not achieve

the same sensitivity as the IPC.

The widest energy range observations of the universal power law spectra of AGN come from the twelve active galaxies that were observed by both the A4 and the A2 instruments on HEAO-1 (Rothschild et al. 1983). These were bright high-latitude x-ray sources selected before flight, consisting mainly of Seyfert I galaxies. With the exception of NGC4151, which showed evidence of photo-electric absorption at low energies, a power law provided a good fit to the combined A2 and A4 spectrum of each source in the 2-165 keV energy range. A thermal bremsstrahlung model was in general less acceptable. The power law spectral indices were distributed about a mean of 1.67 with a statistical error on the mean of 0.03. The indices had a rms deviation of 0.15, setting an approximate upper limit of 0.15 to any intrinsic galaxy-to-galaxy dispersion in spectral index. Rothschild et al. computed the mean spectra of these twelve AGN, which was well fit with a power law with index  $\Gamma = 1.63$ , with no evidence of a break in the spectrum.

Spectra of AGN observed by HEAO-1 A2 in the 2-30 keV range are available for 20 sources (Mushotzky et al. 1980, Mushotzky and Marshall 1980, Mushotzky 1982). These spectra can be described by power laws with differing amounts of low-energy absorption. The mean spectral index is  $\Gamma = 1.65$  with the observed dispersion of 0.1 being consistent within errors of zero intrinsic dispersion.

At the lower energies of Einstein observations, a model more complicated than a simple power law is needed to describe the spectra of AGN. In particular, it is found that some AGN have their spectra altered by photo-electric absorption within the source. Thus the description of the spectra requires a

column density  $n_H$  of absorbing material as well as a power law index.

Petre et al. (1984) derived the spectra of 15 high-luminosity ( $L_{0.75-4.5\text{ keV}} > 10^{43.5}\text{ erg/s}$ ) QSOs and Seyfert galaxies from Einstein SSS data. These spectra showed no evidence of absorbing material intrinsic to the source with  $n_H > 2 \cdot 10^{21}/\text{cm}^2$ . The mean power law index found for the sample was  $\Gamma = 1.66$  while the observed dispersion was consistent with zero intrinsic dispersion. In a subsequent paper Reichert et al. (1985) examined a complementary sample of 12 low-luminosity AGN using Einstein SSS data. They found each spectra to be consistent with a power law with spectral index of  $\Gamma = 1.7$  with one of three cases of absorption: Four of the 12 AGN showed no sign of intrinsic absorption ( $n_H < 10^{21}/\text{cm}^2$ ), three were consistent with a uniform absorbing column, and the remaining five required a 'patchy' absorber that only partially covered the sources. The sources with absorption required mean column densities ranging from  $10^{22}$  to  $2 \cdot 10^{23}/\text{cm}^2$ .

Thus high luminosity AGN show little absorption, while low luminosity AGN appear to separate into two groups; those with little absorption, and those with significant absorption. The absorption is considered to occur in the Broad Line Region, which consists of a swarm of dense clouds each smaller than the central x-ray source. The x-ray absorbing column consists of a small number of clouds, with variations in this number leaving portions of the source totally exposed (Lawrence and Elvis 1982).

It appears then that AGN have a universal spectra, consisting of a power law of spectral index  $\Gamma \approx 1.7$ , at least in the 2 to 165 keV energy range. At energies below 2 keV evidence remains of this universal spectral index. However the spectra are altered by absorption in the source, with the amount of

absorption varying from source to source, and being anti-correlated with the source's x-ray luminosity. The absorption depends on the geometry of material around the source, and as such the AGN spectra in the 0.1-4.5 keV range may be a useful probe of this material. The complicated nature of the absorption however makes the soft x-ray luminosity of an AGN a poor indicator of the luminosity at other energies.

In the low-energy gamma-ray range the spectra of only a few AGN have been measured, with these measurements providing rather loose constraints on the spectral shape above 0.5 MeV. However upper limits from higher energy measurements indicate that AGN spectra must steepen significantly. Bignami et al. (1979) reported the results of searching the SAS-2 data for gamma-ray emission from known x-ray emitting active galaxies. This search resulted in no positive detections, but set upper limits on the 35-100 MeV flux and  $>100$  MeV flux from 30 known x-ray sources. These limits lead to the conclusion that for all the sources studied, the power law spectrum observed at x-ray energies cannot extend into the high-energy gamma-ray range, but rather the spectra must steepen substantially between 50 keV and 35 MeV.

Analysis of the COS-B data supports this conclusion. COS-B detected the QSO 3C 273 (Swanenburg 1978), with a flux in the 50 to 150 MeV range two decades below that predicted from an extrapolation of the x-ray data. A search of the data for gamma-ray emission from known active galaxies produced upper limits for 51 sources, and in some cases these limits were more restrictive than those from SAS-2 (Pollock et al. 1981). Thus the COS-B data reinforces the conclusion that the spectra of AGN in general must steepen between the hard x-ray and high-energy gamma-ray regions.

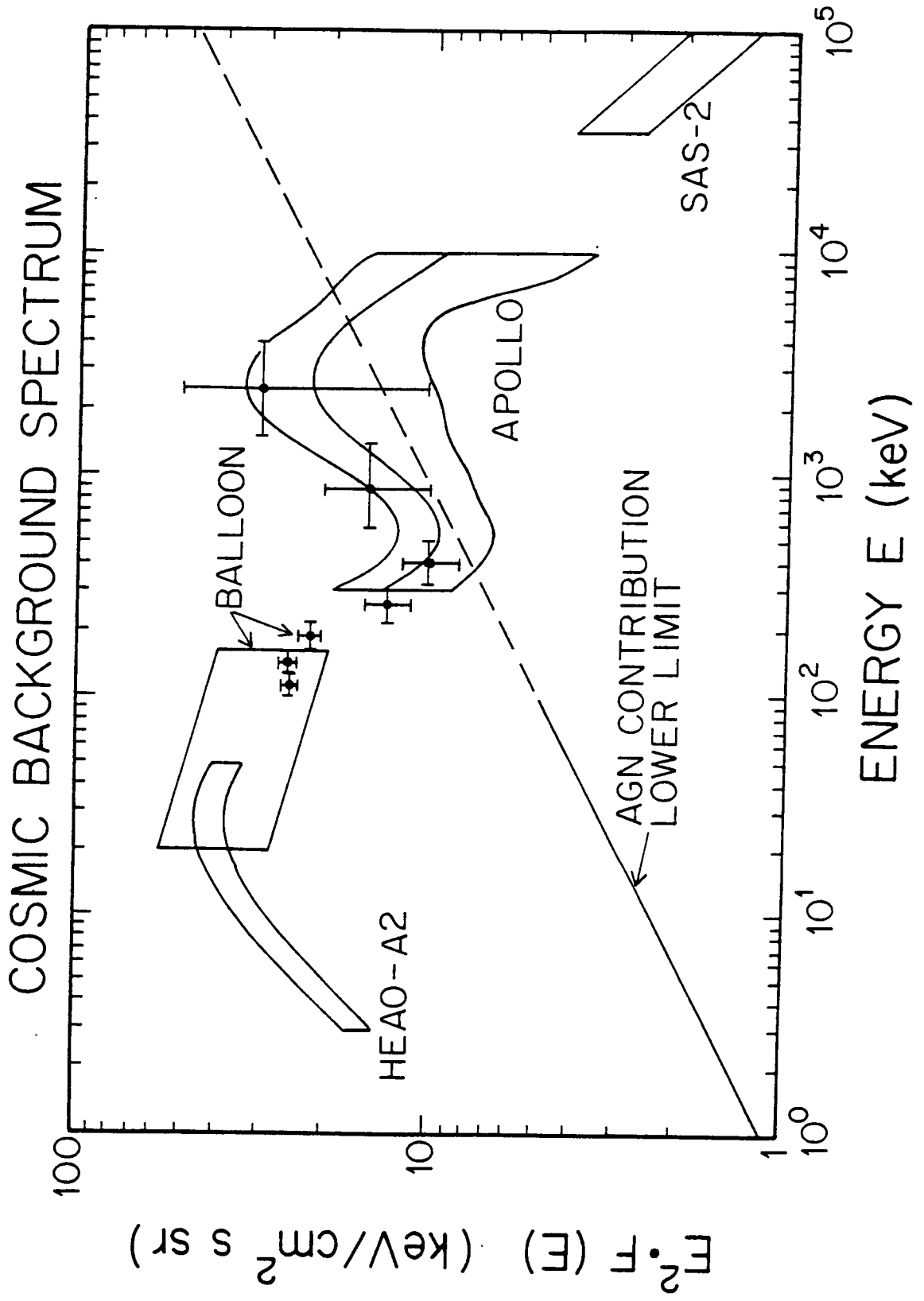
### 1.1.3 The Diffuse X-ray and Gamma-ray Background

The diffuse cosmic background has been a topic of study since the beginning of x-ray and gamma-ray astronomy. This background is due to the combined flux from all extragalactic sources, diffuse intergalactic processes, and any primordial radiation originating from early epochs of the universe. Its interpretation could be cosmologically important, but at present this interpretation is uncertain because of the wide range of sources and processes that could contribute to it. The measured spectrum, however, places a useful upper limit on the total flux received from any single class of sources, constraining both the number and the spectra of these sources.

Figure 1.1 shows the spectrum of the cosmic background from 3 keV to 100 MeV. The Figure shows the energy incident per logarithmic interval of energy. Three measurements from space craft are shown. The data from 3 to 400 keV are from HEAO-1 A-2 (Marshall et al. 1980), while that from 300 keV to 10 MeV are the Apollo results (Trombka et al. 1977), and the spectra above 35 MeV is the SAS-2 measurement (Fichtel et al. 1978). The measurements from two balloon-borne experiments are also shown. These are the 20 to 165 keV results of Kinzer et al. (1978) and the 100 keV to 4 MeV measurements of Fukada et al. (1975). The plot also shows a lower limit to the AGN contribution to the cosmic background, which will be discussed below.

The measurements of the cosmic background are difficult in each energy interval, involving the separation of the cosmic component from other, often dominant, sources of background. For this reason such measurements may suffer from systematic errors. For a discussion of the experimental problems in each energy band see the review by Horstman (1975). In particular the

**Figure 1.1.** The diffuse cosmic background spectrum from 3 keV to 100 MeV. The 3 to 400 keV HEAO-A2 data is from Marshall et al. (1980), the 300 keV to 10 MeV Apollo data is from Trombka et al. (1977), and the SAS-2 data above 35 MeV is from Fichtel et al (1978). The balloon flight results in the 20 to 165 keV interval are from Kinzer et al. (1978), and the balloon flight results from 100 keV to 4 MeV are from Fukada et al. (1975). Uncertainties are shown for the spectral normalization only, except for the Kinzer et al. (1978) data for which point-by-point uncertainties are given, and for the Apollo data, for which the uncertainties are those given by Trombka et al. (1977) for tabulated fit points. Also shown is a lower limit to the active galaxy contribution to the background, which is discussed in the text.



existence of the bump in the spectrum between 1 and 5 MeV has frequently been debated. The measurements of several investigators have however confirmed this feature (White et al. 1977, Schonfelder et al. 1977, Schonfelder et al. 1980).

Historically the cosmic background has been explained as either the superposition of discrete sources, or as due to a diffuse mechanism, with both explanations seeing favor at various times. The 1-5 MeV bump has drawn much attention, with explanations including induced positronium annihilation (Heffernan and Liboff 1984), gravitino decays at high redshift (Olive and Silk 1985),  $^{56}\text{Co}$  decay in supernova, and the superposition of fluxes from Seyfert galaxies with NGC4151 like spectra (Webber et al. 1981).

With the data available from HEAO-1 and HEAO-2, it is now clear that discrete sources can make a significant contribution to the cosmic background at all x-ray and gamma-ray energies. Thus, until the contribution from discrete sources is accurately estimated, the existence of any truly diffuse or primordial component is uncertain. We will discuss the methods for estimating the background contribution due to active galaxies. The first is based on the emissivity per unit volume due to AGN, which, while being more general, produces uncertain results because of the unknown evolution of this emissivity. The second method depends on the surface density of AGNs above a given flux level, and is useful for setting a firm lower limit to the total contribution to the background.

**Emissivity Method.** In a Friedmann cosmology the differential energy flux  $S(E)$  ( $\text{erg}/\text{cm}^2\cdot\text{s}\cdot\text{keV}$ ) received from a source at cosmological redshift  $z$  is given by



$$S(E) = \frac{1}{[1+z]} \frac{L([1+z]E)}{4\pi R^2(z)} \quad (1.1)$$

where  $L(E)$  is the source's differential luminosity (erg/keV·s) as measured in the source's rest frame, and  $4\pi R^2(z)$  is the surface area of the sphere generated by all photons emitted by the source at the same time as those presently being detected. This follows directly from the conservation of photon number, with the factor of  $[1+z]$  arising from the dilation of the time interval between photons. An explicit expression for  $R(z)$  may be found in Misner, Thorne and Wheeler (1973, equation 29.33).

We can describe the spatial and luminosity distribution of a class of sources by a differential luminosity function  $\phi_n(L|E, z)$ , which gives the number of sources per unit luminosity per unit coordinate volume (volume at the present time rather than the time of emission) with differential luminosity  $L$  and energy  $E$  as measured at rest at redshift  $z$ . Partial knowledge of the luminosity function can be obtained from a sky survey, such as the Einstein Medium Sensitivity Survey (Maccacaro et al. 1984) or the HEAO-1 A2 High-Latitude Survey (Piccinotti et al. 1982). The luminosities and redshifts of a sample of sources are in general fit to a parametric model of the luminosity function. At low luminosities and at high redshifts the results will be model-dependent because the survey contains little information about the luminosity function for such sources.

The flux due to this class of sources depends only on the bulk emissivity

$$B_n(E, z) = \int_0^{\infty} \phi_n(L|E, z) L dL \quad (1.2)$$

which gives the total source luminosity per unit coordinate volume. This may have large contributions from low luminosities where the luminosity function is poorly known. In particular, if the luminosity function is described by a power law in a limited range (e.g. Piccinotti et al. 1982) then the bulk emissivity will be very sensitive to the lower luminosity cutoff used. From the bulk emissivity and (1.1) we may calculate the total flux  $S_n(E)$  (erg/cm<sup>2</sup>·s·keV·sr) due to this class of sources as:

$$S_n(E) = \frac{1}{4\pi} \int_0^{z_{\max}} \frac{B_n([1+z]E, z)}{[1+z]4\pi R^2(z)} \frac{dV}{dz} dz \quad (1.3)$$

where  $dV/dz$  is the rate that the total coordinate volume within the sphere defined by redshift  $z$  increases with redshift. We have assumed a cutoff in the luminosity function at a redshift  $z_{\max}$ , which represents the epoch at which galaxy formation began. For a Friedmann cosmology with Hubble constant  $H_0$  and deceleration parameter  $q_0$  (with cosmological constant  $\Lambda=0$ ) the total coordinate volume within redshift  $z$  increases as

$$\frac{dV}{dz} = \frac{4\pi R^2(z)c}{H_0[1+z]\sqrt{1+2q_0z}} \quad (1.4)$$

(Misner, Thorne and Wheeler 1973, equations 29.4 and 29.32a). Therefore the total flux from a class of sources is

$$S_n(E) = \frac{c}{4\pi H_0} \int_0^{z_{\max}} \frac{B_n([1+z]E, z)}{[1+z]^2 \sqrt{1+2q_0z}} dz \quad (1.5)$$

From this result it would appear that the calculation is sensitive to errors in  $H_0$ , and  $q_0$ . However if the value of  $H_0$  used to determine  $\phi_n(L|E, z)$  is

employed, then  $B_n(E,z)/H_o$  is to first order independent of  $H_o$ . The dependence of the result on  $q_o$  is small (Avni 1978) compared to the uncertainties introduced by the unknown behavior of  $\phi_n(L|E,z)$  at small luminosities and large redshifts.

Several authors have attempted these calculations, producing various estimates of the contribution of active galaxies to the cosmic background. From the Einstein Medium Sensitivity Survey, Maccacaro et al. (1984) estimated that 62-123% of the 2 keV background is due to active galaxies. They assume the luminosity function evolved by pure luminosity evolution, which is equivalent to assuming that the luminosities of individual sources decrease exponentially in time, with the present distribution in luminosity being homogeneous in space. For pure luminosity evolution the redshift dependence of the luminosity function is

$$\phi_n(L|E,z) = \int_0^{\infty} \phi_n(L_o|E,0) \delta(L - e^{C\tau(z)} L_o) dL_o \quad (1.6)$$

where  $\tau(z)$  is the look-back time, the time in units of  $1/H_o$  since presently detected photons were emitted by a source at redshift  $z$ , and the evolutionary parameter  $C$  is obtained by model fitting. The evolutionary parameter determined by Maccacaro et al. (1984) was  $C = 4.85$  with a 95% confidence interval of  $[2.94, 6.11]$ , while the  $z=0$  luminosity function determined fit a power law  $\phi \propto L^{-\gamma}$  with  $\gamma = 3.6$  for luminosities above  $10^{25}$  erg/s·Hz at 2 keV, but falling below this relationship for lower luminosities. This turnover in the luminosity function may be due to absorption in low luminosity sources.

Piccinotti et al. (1982) estimated from the HEAO-1 A2 data that about 20% of the 2-10 keV background is due to AGN. The HEAO-1 A2 High-Latitude Survey contained only nearby sources with the largest AGN redshift being 0.158, so that only a local luminosity function could be derived. The 20% estimate therefore assumes no evolution of the luminosity function.

Cavaliere et al. (1985) suggest a differential luminosity evolution model for the luminosity function, in which evolution is strongest for high-luminosity sources, and does not occur at all for the lowest luminosity sources. This model is based on the conception that AGN are fueled by a compact mass stockpile that is depleted because of insufficient replenishment from the host galaxy plus a steady mass inflow largely independent of the compact mass supply. Based on this model, Danese et al. (1986) have computed luminosity functions compatible with the HEAO-1 A2 data and the Einstein Medium Sensitivity Survey data that predict 23% to 53% of the 2 to 10 keV background being contributed by active galaxies. These results indicate then that a sizeable fraction of the x-ray background could be due to the flux from active galaxies. Accurate estimation of this contribution, however, must wait for a more accurate determination (perhaps by AXAF) of the AGN luminosity function's low luminosity and evolutionary behavior.

**Surface Density Method.** There is an alternate formulation for determining the background contribution of a class of sources that allows us to set firm lower limits. The flux from a class of sources is given by

$$S_n(E) = \int N_n(S) S dS \quad (1.7)$$

where  $N_n(S)$  is the surface density of sources in this class (number per solid

angle per unit flux) with source flux  $S$ . This places the burden of the calculation upon the estimation of  $N_n(S)$  which for low fluxes must be calculated from the poorly known luminosity function with assumptions about cosmology (see Section 1.3). For large source fluxes however  $N_n(S)$  is directly measurable, allowing us to place accurate lower limits on the discrete source contribution to the background.

Survey	Energy	Relation
MSS	0.3-3.5 keV	$N_{AGN}( > S ) = \frac{3530}{sr} \left( \frac{S}{2 \cdot 10^{-13} \text{ erg/cm}^2 \cdot s} \right)^{-1.71 \pm 0.15}$ $N_{clusters}( > S ) = \frac{1100}{sr} \left( \frac{S}{2 \cdot 10^{-13} \text{ erg/cm}^2 \cdot s} \right)^{-1.04 \pm 0.23}$
HSS	1-3 keV	$N( > 2.6 \cdot 10^{-14} \text{ erg/cm}^2 \cdot s ) = (6.3 \pm 2.8) \cdot 10^4 / sr$

**Table 1.2.** Einstein Observatory surface density results. The Medium Sensitivity Survey results are from Gioia et al. (1984) and apply to 0.3-3.5 keV fluxes in the  $7 \cdot 10^{-14}$ - $2 \cdot 10^{-12}$  erg/cm<sup>2</sup>·s range. The High Sensitivity Survey result is from Giacconi et al. (1979b) and applies to the 1-3 keV flux of all extragalactic sources.

Table 1.2 summarizes the Einstein Observatory surface density measurements. From the Einstein Medium Sensitivity Survey Gioia et al. (1984) have determined  $N_n(S)$  separately for AGN and galactic clusters. For a Euclidean universe with no expansion the  $N_n(S)$  should be proportional to  $S^{-2.5}$ . They found instead a dependence on  $S^{-2.71}$  for AGN, and a dependence on  $S^{-1.04}$  for

galactic clusters. This is direct evidence for evolution with more luminous or numerous AGN in the past, and less luminous or numerous clusters in the past. The Einstein High Sensitivity Survey (Giacconi et al. 1979b) measured  $N(>S)$  for extragalactic objects at a 1-3 keV flux of  $2.6 \cdot 10^{-14} \text{ erg/cm}^2 \cdot \text{s}$  (or  $4.5 \cdot 10^{-14} \text{ erg/cm}^2 \cdot \text{s}$  for 0.3-3.5 keV assuming a power law  $\Gamma=1.67$ ). This result is compatible with the Medium Sensitivity Survey AGN surface density relation. Giacconi et al. (1979b) estimated the contribution from  $S > 2.6 \cdot 10^{-14} \text{ erg/s}$  (1-3 keV) as  $26 \pm 11\%$  of the background, assuming  $N(S) \propto S^{-2.5}$ . If we instead use the Medium Sensitivity Survey results integrated over the range  $4.5 \cdot 10^{-14} - 10^{-11} \text{ erg/cm}^2$ , we obtain lower limits to the discrete source contributions of 10% of the background for AGN and 3% of the background for clusters at 2 keV (where we have extrapolated the 2 keV background level from the Marshall et al. (1980) measurement).

The galaxy clusters have a thermal spectrum which falls quickly with increasing energy; however, as we have seen, the active galaxies have a rather hard spectrum, which will increase their importance at higher energies. If we assume that all AGN have a spectra with the hard x-ray power law index of  $\Gamma = 1.67$ , then we can extend the lower limit to other portions of the spectrum. This is shown in Figure 1.1. In this extrapolation no correction has been made for the obscuration observed in the Einstein energy range. Such a correction would tend to increase the estimate of the AGN contribution at higher energies.

Clearly, as several authors have pointed out (e.g. Rothschild et al. 1983, Bassani and Dean 1984) to avoid over-subscribing the background, a substantial fraction of AGN spectra must turn over below an observed energy of 5

MeV, where the AGN lower limit and the background spectra meet. We may also conclude that if a large fraction of AGN spectra extend to the MeV region, then the background above 0.5 MeV is dominated by the AGN contribution. If the AGN contribution is as large as 50% or larger in the soft x-ray as suggested by Maccacaro et al.(1984), then the spectra of AGN must turn over in the hard x-ray, which has not been seen. This suggests that the pure luminosity evolution model that they have used to fit the Medium Sensitivity Survey is at odds with the existing hard x-ray data.

## 1.2 Source Models

The common features of active galaxies, such as their compact, highly luminous nature, their emission line structure, and the universal nature of their hard x-ray spectra, have engendered the hope that one day these objects may be explained by a single model. The bewildering diversity of optical and radio behavior which has led to numerous classes and subclasses of AGN would then be seen as secondary characteristics that hide the intrinsic similarity of these objects. The success of such a unification will depend on wide-band observations of many sources, with a key role to be played by gamma-ray astronomers. At present, while tentative attempts have been made at constructing a unified picture of AGN (Begelman 1986, Blandford 1985), no widely accepted general theory exists. We must therefore be satisfied with a few unifying concepts.

**Compactness.** A feature common to all AGN is a large luminosity ( $10^{40}$ - $10^{47}$  erg/s) emanating from a relatively small region ( $10^{12}$ - $10^{17}$  cm). This compactness has motivated many models of AGN that are based on a central

massive ( $10^6$ - $10^{10} M_{\odot}$ ) black hole. For a review of these models see Blandford (1985), and Rees (1984). We will consider here only the evidence for compact central sources, and a few of the consequences of this compactness.

Direct evidence for the compactness of AGN comes from the comparison of their luminosities with their minimum variability time-scales. Figure 1.2 compares the 2-10 keV luminosities of several AGN with their minimum x-ray variability time-scales. The data is from the compilations of Barr and Mushotzky (1986), and Wandel and Yahil (1985). The variability time-scale  $\Delta t$  is the minimum time period over which the source flux changes by a factor of two. Only Seyferts and QSOs for which this time-scale is resolved within a factor of two are shown. For the BL Lacs shown, the variability time-scales are less well-resolved, and may in general be taken only as upper limits.

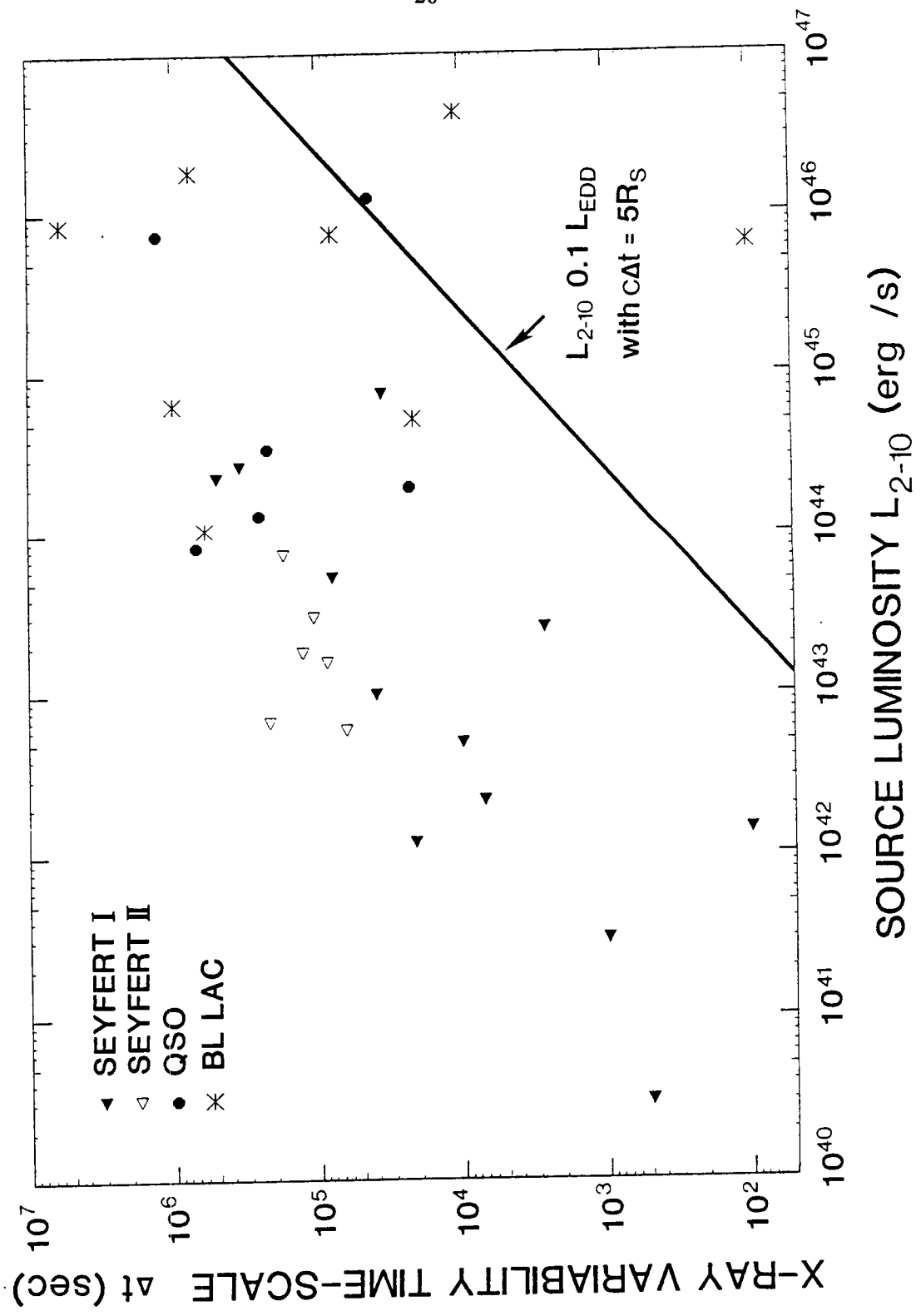
The minimum variability time-scale  $\Delta t$  of a source provides an upper limit to the radius  $R_s$  of the x-ray source region. We expect  $c\Delta t = \alpha R_s$  where  $c$  is the speed of light. The proportionality constant  $\alpha$  is model-dependent, but always greater than unity. The only exception to this is if the variable flux is due to material in bulk relativistic motion. In the context of a black hole model, it is of interest to compare this radius to the source's Schwarzschild radius

$$R_s = \frac{2GM}{c^2} = 2.98 \cdot 10^5 \text{ cm} \left( \frac{M}{M_{\odot}} \right) < c\Delta t \quad (1.8)$$

where  $G$  is Newton's constant. This comparison produces an upper limit to the source's mass.



**Figure 1.2.** Minimum x-ray variability time scale versus 2-10 keV luminosity for active galaxies. The time-scale  $\Delta t$  is the minimum two-folding time, which for the Seyferts and QSOs shown has been resolved within a factor of two. The data is from the tabulations of Barr and Mushotzky (1986), and Wandel and Yahil (1985). The line shows the relations between x-ray luminosity and variability time scale for Eddington limited sources which radiate 10% of their luminosity in the 2 to 10 keV band, assuming a variability time scale of five times their Schwarzschild radii over the speed of light.



In contrast, the luminosity of a source may be used to determine a lower limit to its mass. If a source is assumed to be gravitationally bound, then the balance of gravity with radiation pressure at the surface of the radiation's last scatter sets an upper limit to the source's luminosity for a given source mass. This limiting luminosity, the Eddington luminosity, is given by

$$L_{EDD} = \frac{4\pi c G M m_p}{\sigma_T} = 1.26 \cdot 10^{38} \text{ erg/s} \left( \frac{M}{M_\odot} \right) \quad (1.9)$$

where  $M$  is the source mass,  $m_p$  the proton mass, and  $\sigma_T$  the Thomson cross section. The derivation of this limit assumes a spherically symmetric distribution of fully ionized hydrogen plasma around an isotropic central source, with the bulk of the radiation of energy lower than the electron rest mass. The Eddington luminosity provides reasonable limit to the order of magnitude of the source luminosity even when these conditions are not met, with the notable exception being the condition of the source's isotropy. If the radiation is beamed, then the luminosity inferred for an isotropic source from the observed flux can greatly exceed the Eddington limit. The lower limits to source masses set by requiring the 2-10 keV x-ray luminosity of AGN to be less than their Eddington luminosity, range from  $\sim 10^2 M_\odot$  for low luminosity Seyfert I galaxies, to  $\sim 10^9 M_\odot$  for high luminosity QSOs.

If we assume the presence of a central black hole, with the x-ray flux emanating from a radius of  $R_* = 5R_g$  (Lightman et al. 1979), and require that the total luminosity be less than the Eddington limit, we obtain a limiting relationship between  $L$ , the source's total luminosity, and  $\Delta t$ , the source's minimum variability time-scale. This is shown in Figure 1.2 where we have

assumed the bolometric correction  $L = 10L_{2-10}$ . This correction is rather uncertain and assumes the source spectra continue out to 1 MeV with a power law slope of  $\Gamma = 1.67$ . Gamma-ray observations would result in improved luminosities, and perhaps shorter variability time-scales. The Seyferts and most of the QSOs have luminosities below this limit, and in this respect are compatible with black holes radiating at  $10^{-1}$  to  $10^{-3}$  of their Eddington limit. Two of the BL Lacs and one of the QSOs violate this limit, suggesting that the sources are beamed.

**The Energy Source.** If the central source is indeed a massive black hole, then the energy needed to create the observed luminosity may be supplied by mass accretion or be extracted from the rotational energy of the black hole by the Blandford-Znajek mechanism. The Blandford-Znajek mechanism (Blandford and Znajek 1977) requires only a very small accretion rate, but implies a former era of accretion to build up a massive rotating black hole. If the source is powered by accretion then the accretion rate required is

$$\dot{M} = 0.176 M_{\odot}/yr \frac{0.1}{\eta} \frac{L}{10^{45} \text{ erg/s}} \quad (1.10)$$

where  $\eta$  is the efficiency with which mass is converted to energy (typically estimated as 0.1). This gives the rate of growth of the black hole mass. The exponential time constant associated with this rate of growth is  $\tau = M/\dot{M}$ , which, relative to the Hubble time  $1/H_0$ , is

$$H_0 \tau = \frac{\eta}{40} \left( \frac{L_{EDD}}{L} \right) \quad (1.11)$$

assuming a value of 55 km/s-Mpc for  $H_0$ . Thus in order for a massive black hole to form by accretion in a time short compared to the Hubble time, either it must experience a period of low efficiency (e.g. swallowing stars whole), or it must radiate on average at a reasonable fraction ( $\sim .01$ ) of its Eddington luminosity.

There are several sources possible for the accreting material. These include the interstellar medium, gas from tidally disrupted stars, and gas loss from stars as a result of stellar collisions. The latter two sources present some interesting possibilities. The stellar disruption and collision rates could be greatly enhanced by interactions between galaxies. Whether this is actually the case is at present debatable, due to our poor understanding of the transport of matter to the AGN from the surrounding galaxy. If, as suggested by Begelman Blandford and Rees (1984), all galaxies contain a central massive black hole, then, as suggested by Roos (1985a), conditions in AGN may differ from those in normal galaxies only because of recent mergers or interactions. While little empirical evidence for this view exists, QSOs do appear to be associated with regions of higher-than-average galactic density (Stockton, 1985). IRAS infrared observations of interacting pairs of galaxies have shown that signs of nuclear activity are more common in interacting galaxies than non-interacting galaxies (Cutri and McAlary, 1985). In particular a population with extremely luminous infrared emission was found to be unique to interacting pairs. If galaxy interactions are important in initiating nuclear activity, the luminosity distribution of AGN and its evolution might be explained on the basis of merger and interaction rates (Roos 1985a,b).

**Radiation Mechanisms.** Continuum x-rays and gamma-rays from astrophysical sources are principally the result of high-energy electrons. The high-energy photons may be either created directly by bremsstrahlung or synchrotron radiation, or indirectly by the inverse Compton scattering of low-energy photons. In a high-temperature thermal plasma the important processes are thermal bremsstrahlung and Comptonization, with the latter being the dominant source of electron cooling if the photon number density exceeds the particle number density. For non-thermal (i.e. power law) relativistic electrons the important processes are synchrotron radiation and inverse Compton scattering, Comptonization dominating if the radiant energy density exceeds the magnetic field energy density (Fabian 1984).

A popular combination of these processes is the Synchrotron Self Compton model (SSC). In this model it is assumed that a major fraction of the nuclear radio to ultraviolet flux is synchrotron radiation, produced by relativistic electrons in a magnetic field. The synchrotron photons then inverse Compton scatter on the relativistic electrons to produce the x-ray and gamma-ray spectrum. This model predicts a two-component spectrum, consisting of the initial synchrotron spectrum and a Compton scattered spectrum which mimics the form of the synchrotron spectrum but is shifted to higher energies. Flux variation in the two components of the spectrum are expected to be correlated, so that testing of the model must rely on simultaneous broad-band measurements. The power law nature of the hard x-ray spectrum is in this model a result of the initial power law electron spectrum, but at present there is no explanation for the universal nature of its slope. For applications of the SSC model to AGN see Bassini (1981), Urry (1984), and

Band and Grindlay (1986).

Many authors have suggested that pair creation processes are important in AGN. Electron-positron pairs are created by the interaction of two photons with a total center of momentum energy greater than  $2m_e c^2 \approx 1.0$  MeV. Sources with high photon density can therefore be optically thick to pair creation for photons with energy above this threshold. Guilbert et al. (1983) have shown that a source with a primary radiation mechanism which produces (before considering pair production) a luminosity  $L_\gamma$  above 1.0 MeV, will be thick to pair production if

$$\left( \frac{R_s}{R} \right) \left( \frac{L_\gamma}{L_{EDD}} \right) > \frac{m_e}{m_p} = \frac{1}{1835} . \quad (1.12)$$

If this is the case, then a electron-positron plasma is generated, and the emergent gamma-ray spectra is cutoff above  $\sim 1.0$  MeV. The pair plasma generates lower-energy photons, and Comptonizes the lower-energy portion of the spectrum. Several author have attempted to calculate the resultant spectrum under various assumptions; for examples see Fabian (1985), Zdziarski and Lightman (1985), and Zdziarski (1984).

**Jets.** There is evidence that the x-radiation from some AGN is beamed. As we have seen, the variability of some BL Lacs and QSOs require beaming to explain their flux variability. Bassani and Dean (1986) have estimated the attenuation of high-energy photons due to pair production for several sources, and found that the positive 50 MeV detection of 3C273 can only be explained by invoking beaming. Perhaps then it is no coincidence that 3C273 is also a superluminal radio source, as superluminal motion is commonly interpreted as

a jet-like outflow at relativistic speed oriented near our viewing direction.

If x-ray and gamma-ray producing jets are common in AGN, then flux limited surveys will be strongly biased to include objects that contain jets oriented in our viewing direction. Understanding which objects are beamed and which are not then becomes an important consideration in modeling AGN. Correlation with radio features may be important in this respect, but direct determination of jet parameters from gamma-ray observations should also play an important role.

### 1.3 Future Expectations

Future observations of active galaxies in low-energy gamma-rays should provide information important in understanding their nature and evolution. The observations needed are of two kinds. Detailed observation of the spectrum and variability of individual sources are needed to provide a general characterization of the gamma-ray emission of AGN. Flux limited gamma-ray sky surveys are needed for statistical studies of the distribution and evolution of the AGN population. For either form of observation, progress in understanding AGN will require instruments with improved sensitivity. This improved sensitivity, perhaps obtained by more sophisticated background rejection and the prolonged exposures available from satellites, will allow more detailed source studies and an increase in the number of detectable sources.

With twenty or so AGN observed in low-energy gamma-rays, we will know whether AGN spectra have a universal form or vary widely from source to source. If the universal x-ray power law continues out to the MeV region



with a turnover at an energy consistent from source to source, then mechanisms that produce such a spectrum under a wide range of conditions must be sought. In this case the measured cosmic background in the MeV region will tightly constrain the evolution of the AGN luminosity function and its behavior at low luminosities. If however AGN have no universal gamma-ray spectra, then AGN gamma-ray properties will provide clues to the difference between the central sources of the various AGN classes.

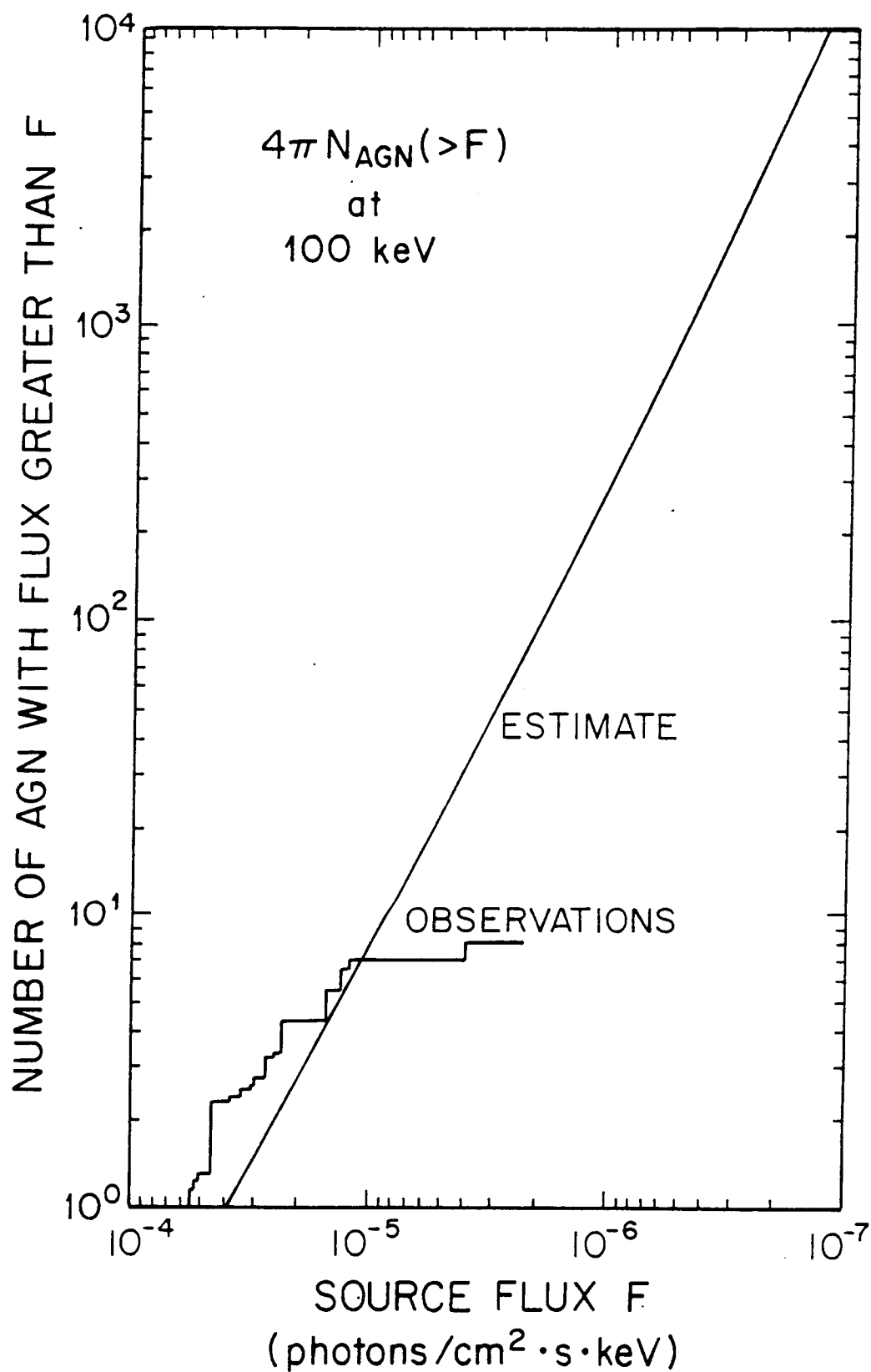
With one hundred or so AGN observed in low-energy gamma-rays, detailed statistical studies will begin to reveal their distribution in redshift and luminosity. The ultimate goal of such studies is to determine the life course of individual sources, and the history of the total source population. These studies will also improve the understanding of the diffuse background.

### 1.3.1 Estimates of $N(>F)$

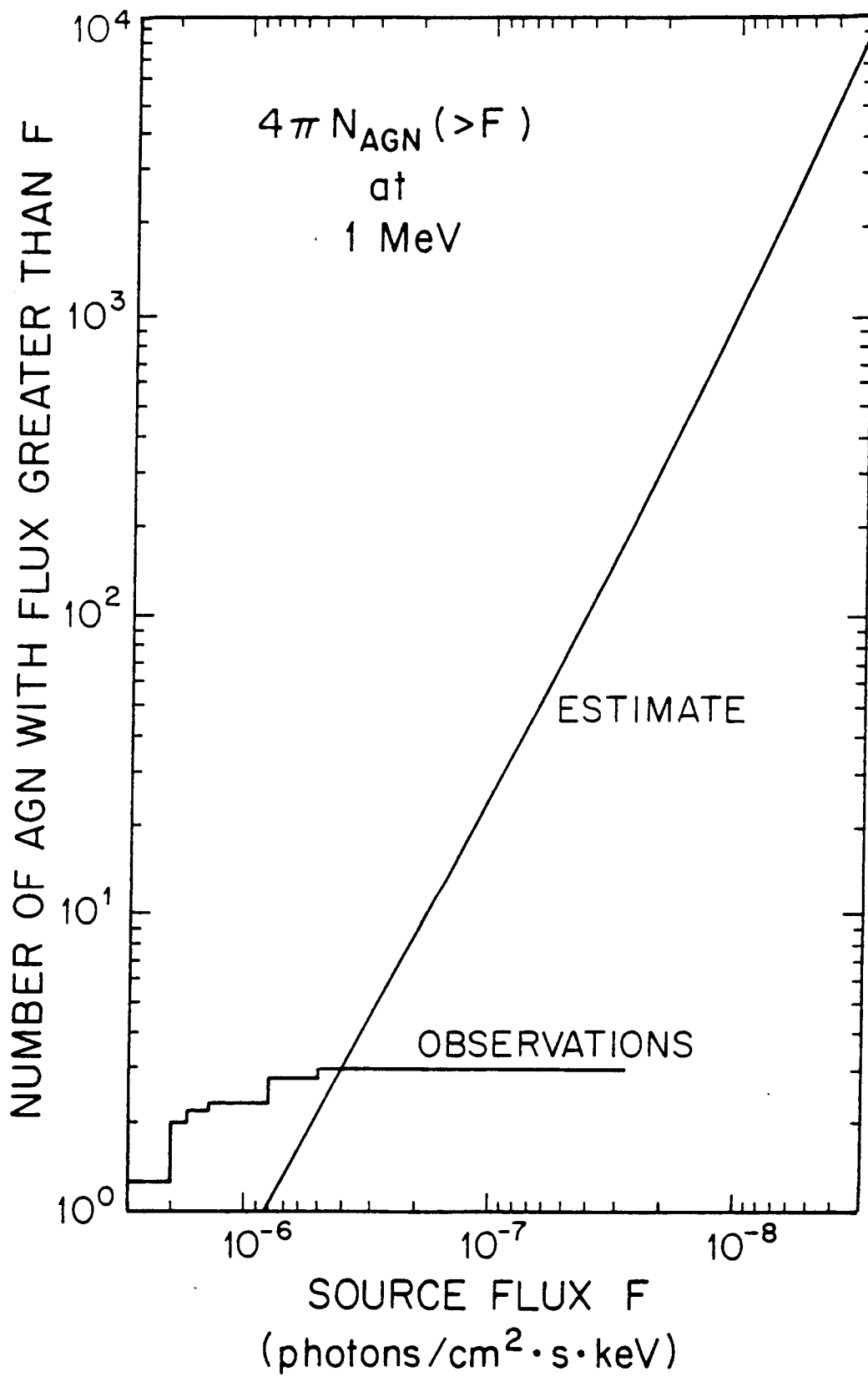
The scientific output of observational low-energy gamma-ray astronomy will depend on the number of sources that are detectable. An experimentally important question then is: What sensitivity is needed to detect a given number of sources? In this section we will consider the answer to this question, with estimates for AGN of  $N(>F)$ , the surface density of sources with number flux at a given energy above the sensitivity  $F$ .

We will estimate  $N_{AGN}(>F)$  at 100 keV and 1 MeV. The energy of 100 keV was chosen because the power law nature of AGN spectra is relatively well-established at this energy, while 1 MeV was chosen to be representative of the low-energy gamma-ray band. These estimates are given in Figures 1.3 and 1.4 respectively, which show  $4\pi N_{AGN}(>F)$ , the number of AGN detectable

**Figure 1.3.** An estimate of the number of active galaxies observable in the full sky at 100 keV versus instrument sensitivity. Also shown is the mean number of AGN observed to date versus sensitivity.



**Figure 1.4.** An estimate of the number of active galaxies observable in the full sky at 1 MeV versus instrument sensitivity. Also shown is the mean number of AGN observed to date versus sensitivity.



Source	Type	Energy	Flux ( $10^{-8}/\text{cm}^2 \cdot \text{s} \cdot \text{keV}$ )	Reference
3C390.3	Seyfert I	100 keV	23	12
NGC1275	Seyfert I	100 keV	12	12
			27	3
Mkn279	Seyfert I	100 keV	39	12
3C273	QSO	100 keV	15	9
NCG6814	Seyfert I	100 keV	13	12
MGC8-11-11	Seyfert I	100 keV	45	3
		1 MeV	6.5	
NGC4151	Seyfert I	100 keV	34	12
			15,25,30,31,34, 38,51,53,63,65,74	2
		1 MeV	<0.5	14
			<1.7,1.8	2
			<2	13
			2	11
			5	8
			<12	5
			14	10
CEN A	Radio galaxy	100 keV	15,55,55,60,85,100, 135,140,140,140,135, 155,205	4
		1 MeV	0.8	1
			<2	6
			2	7

**Table 1.3.** Active galaxies detected at 100 keV or 1 MeV. For each source the table lists fluxes at 100 keV reported in the literature. For those sources detected at 1 MeV the table lists reported fluxes and upper limits. Fluxes and upper limits are from the following sources:

- |                           |                              |
|---------------------------|------------------------------|
| 1) Baity et al. (1981)    | 2) Baity et al. (1984)       |
| 3) Bassani et al. (1981)  | 4) Bealle et al. (1978)      |
| 5) Coe et al. (1981)      | 6) Gerhals et al. (1984)     |
| 7) Hall et al. (1976)     | 8) Megan and Haymes (1979)   |
| 9) Primini et al. (1979)  | 10) Perotti et al. (1979)    |
| 11) Perotti et al. (1981) | 12) Rothschild et al. (1983) |
| 13) Schonfelder (1980)    | 14) White et al. (1980)      |

in the full sky, versus the source flux limit  $F$  in units of photons/cm<sup>2</sup>·s·keV. At the largest source fluxes we may use detections reported in the literature as a lower limit to the number of detectable sources. These are listed in table 1.3 and in Figures 1.3 and 1.4 represented by the stair-stepped curves, which give the cumulative number of sources detected above a source flux  $F$ . For sources with multiple detections or reported upper limits, the ratio of the number of detections above a given flux, to the number of upper-limits below this flux plus the total number of detections was computed. This ratio for all the sources was then summed.

To obtain estimates at lower flux levels we must extrapolate from the more numerous observations in other energy bands. We will employ the x-ray data and assume that each source has a power law spectra with index  $\Gamma = 1.67$  over all relevant energies. For extrapolation to 100 keV this is uncontroversial, but for extrapolation to 1 MeV we are assuming that the spectral break is beyond 1 MeV for all AGN. With the data available (1 out of 3 sources has shown a break in the spectrum, this being at  $\sim 3$  MeV) we can proceed in no other fashion. However if this assumption is not true, our estimate will need to be reduced by the fraction of AGN with spectral breaks below 1 MeV.

The x-ray data to be considered consists of the  $N_{AGN}( > S )$  relation determined for the Medium Sensitivity Survey, and the local luminosity function determined from the HEAO-1 A2 High Latitude Survey. The Einstein survey provides the deepest sampling in redshift, with the evolution of the luminosity function apparent in  $N_{AGN}( > S )$ . Unfortunately the  $N( > S )$  predicted for the HEAO-1 A2 survey from the Einstein data is approximately a factor of three

lower than that measured. This prediction involves the extrapolation from the Einstein 0.3-3.5 keV energy range to the 2-10 keV HEAO-1 A2 range. The two surveys also disagree on the shape of the local luminosity function. As pointed out by Maccacaro et al. (1984), photo-electric absorption which is found mainly in low-luminosity sources, causes low luminosity sources to be under-represented in the Medium Sensitivity Survey. Reichert et al. (1985) showed that this obscuration can qualitatively explain the difference between the local luminosity functions. Thus because of the obscuration present in the Einstein energy band, we cannot use the Medium Sensitivity Survey results directly.

Piccinotti et al. (1982) obtained a local luminosity function for AGN from the HEAO-1 A2 survey. This is a power law in the source luminosity to the  $-2.75$  power over the 2-10 keV luminosity range of  $3 \cdot 10^{42}$  to  $1.5 \cdot 10^{45}$  erg/s. Near the bottom of this range there are indications of a turnover in the luminosity function. For lower luminosities Elvis et al. (1984) have estimated the luminosity function based on optical data and optical x-ray correlations. To obtain an  $N_{AGN}( > S )$  relation from this luminosity function, we need a cosmological model and a model of how the luminosity function evolves with redshift. For high source fluxes the resulting surface densities are model independent and may in fact be calculated from a simple Euclidean model. This simple model however does not incorporate the evolution evident in the Medium Sensitivity Survey or comparable optical surveys. We will therefore use the  $q_0=0$  cosmology and the pure luminosity evolution model employed by Gioia et al. (1984) in their analysis of the Medium Sensitivity Survey. In this model the evolution of the luminosity function is given by



(1.6) where the evolutionary constant  $C = 4.85$  and the look back time takes on the simple form  $\tau = z/(1+z)$ . With the source flux in units of ( $\text{erg}/\text{cm}^2 \cdot \text{s} \cdot \text{keV}$ ) and luminosities in units of ( $\text{erg}/\text{s} \cdot \text{keV}$ ) we may calculate the AGN surface density from

$$N_{AGN}(S) = \frac{1}{4\pi} \int_0^{\infty} \int_0^{\infty} \phi_{AGN}(L|[1+z]E, z) \delta\left(S - \frac{L}{4\pi[1+z]R^2(z)}\right) dL \frac{dV}{dz} dz \quad (1.13)$$

The number flux relation  $N_{AGN}( > F )$  resulting from this model is shown extrapolated to 100 keV and 1 MeV in Figures 1.2 and 1.3. For large fluxes the log-log slope is 1.5 as would be found with a Euclidean calculation. At lower flux levels the slope increases, reaching a value of 1.75 at a flux of  $10^{-7}$  photons/ $\text{cm}^2 \cdot \text{s} \cdot \text{keV}$  for the 100 keV curve. The curve for 100 keV is in fair agreement at large fluxes with the reported detections. The relation at 1 MeV however greatly under-predicts the current number of detections. For the given  $N_{AGN}( > F )$  relation, the probability of detecting three or more AGN above a flux of  $1.5 \cdot 10^{-6}$  photons/ $\text{cm}^2 \cdot \text{s} \cdot \text{keV}$  is only 0.007. Thus either we happen to be favorably situated near these sources, or our assumptions about the AGN spectra that were used to extrapolate from 100 keV to 1 MeV are wrong. The predicted number of sources is fairly sensitive to these assumptions. We have assumed that there is no dispersion in the spectra indices, if the dispersion is as large as the 0.15 upper limit of Rothschild et al. (1983) then  $N_{AGN}( > F )$  will be larger by a factor of nearly two. This is due to the bias toward detecting the sources with harder spectra. We could also consider alternate forms for the spectra above 100 keV. However any assumed spectral form that increases  $N_{AGN}( > F )$  at 1 MeV will also increase the AGN

contribution to the background at 1 MeV. To avoid over-subscribing this background and yet increase the surface density at high fluxes at the same time, we would need to postulate a harder spectrum extending out beyond 1 MeV for nearby sources, but with distant sources having a break in their spectrum below 0.5 MeV. At present, the evidence that at 1 MeV  $N_{AGN}( > F )$  is higher than our estimate, is fairly weak; however it is clear that only a handful of future detections establishing the present experimental trend would provide evidence for spectral evolution of AGN.

### 1.3.2 Instrumental Requirements

Important contributions to the understanding of active galaxies can be made by instruments operating in the 100 keV to 10 MeV region. If such an instrument had a sensitivity of  $2 \cdot 10^{-6}$  photons/cm<sup>2</sup>·s·keV at 100 keV and  $4 \cdot 10^{-8}$  photons/cm<sup>2</sup>·s·keV at 1 MeV it would be capable of observing one hundred AGN and observing the brightest AGN in great detail.

To obtain this sensitivity an instrument must be provided with sufficient background suppression, detector area and observation time to reduce the statistical uncertainty on a flux measurement to a fraction of this level. Just as important as the statistical uncertainty of a flux measurement however is the systematic uncertainty. Gamma-ray instruments operate in environments with backgrounds much larger than the astrophysical fluxes being measured. Small fractional errors in the measurement of these backgrounds can have major impacts on the flux measurements. Observational techniques that require physical motions or changes of the instrument, or prolonged periods of time between source and background measurements are subject to systematic

errors due to slight alterations of the background.

In addition, an instrument with the statistical and systematic capability to detect 100 sources will in fact not be able to do so if it cannot individually resolve them. To resolve 90% of these 100 sources from each other requires at a minimum an angular resolution of 3.6 degrees. However in reality this angular resolution is not sufficient. The sky contains numerous weak sources that cannot be individually detected, yet the total flux from such sources within an instrument's field-of-view can vary significantly with direction, and is in fact a source of systematic error in any source flux measurement. Consideration of the angular resolution requirements of a gamma-ray instrument will be the topic of the next chapter.

## **2. Angular Resolution Requirements for Gamma-ray Instruments**

The past achievements of gamma-ray astronomy have been accomplished with instruments having fairly modest angular resolution. These instruments have typically been collimated detectors with active shielding and a field-of-view ranging from 10 to 1000 square degrees, or Compton telescopes with an angular resolution on the order of 5 degrees. Future instruments, capable of detecting tens or hundreds of sources, will need much finer angular resolution. The need for this resolution is driven by two factors, the need to establish source locations with sufficient accuracy for determining unique counterparts at other energies, and the need to reduce the confusion of source flux measurements caused by other nearby sources. In this chapter we will consider the angular source localization requirements for successful counterpart identification, and the angular resolution requirements imposed by source confusion considerations.

### **2.1 Source Localization**

Basic to the study of any gamma-ray source is the determination of its source direction. Accurate determination of this direction is needed for the comparison of measurements at different times or different energy bands. For newly-detected sources it is particularly important to establish a source's direction with sufficient accuracy to determine unique x-ray and optical

counterparts. These identifications allow wide band studies of individual sources, and comparison of properties of source populations at different energies. We will assume that it is sufficient in most cases to be capable of identifying unique x-ray counterparts. The more accurate x-ray source direction can then be used for identifying counterparts in other energy bands.

The determination of a counterpart of a gamma-ray source in the x-ray energy band requires searching within a plausible region of the sky for a source of plausible type with flux within a plausible range. If such a source is found, and the random chance of finding such a source is small, then we have succeeded in identifying a counterpart. The plausible region of the sky is determined by the measured gamma-ray source direction and its error, which we will call the localization accuracy. We wish to make this small enough that the random chance of finding a plausible counterpart is also small. We will assume the counterpart is an active galaxy. Thus to determine the desired localization accuracy, we need to determine a plausible range of x-ray source fluxes, and the surface density of AGNs with these fluxes.

To be specific let us consider an instrument which operates from 100 keV to several MeV. The spectrum measured for a source will in general be best determined at lower energies, due to the nature of gamma-ray backgrounds. Let us suppose then that the instrument has detected a source with flux  $F$  such that the surface density of AGNs (number per solid angle) at 100 keV is  $N_{AGN}( > F ) = N$ . From the 100 keV flux we can extrapolate to the 2 keV x-ray flux  $S_x$  using the known hard x-ray spectral shape. We may then search x-ray catalogues for sources in the same region with reported flux near this extrapolated value. We will find a surface density in the x-ray of

$N_{AGN}( > S_x ) \approx N$ . However, AGN are variable, and the extrapolation poor in the soft x-ray, so that it is plausible that any existing measurement of the sources x-ray flux is lower than the extrapolation by as much as a factor of 10. Taking  $N_{AGN}( > S_x )$  as a power law with index -1.71, we can thus estimate the surface density of plausible counterparts as  $50N$ . We will assume that the random chance of a plausible source being within the area  $\pi\sigma_\theta^2$  is less than .1, where  $\sigma_\theta$  is the localization accuracy. This requires a location accuracy of

$$\sigma_\theta < \frac{.025}{\sqrt{N}} \quad (2.1)$$

or less than a half degree if 100 sources are detectable in the full sky.

## 2.2 Source Confusion

The  $N(>F)$  relation allows us to predict the number of sources that will be observable with an instrument of a given sensitivity. However if the instrument does not have sufficient angular resolution, observation of these sources will be difficult if not impossible. With poor angular resolution, confusion of the weaker sources will hamper the instrument's ability to measure their fluxes. Source confusion is a condition where the flux measurement of a source is seriously affected by the presence of other nearby sources. In the simplest case the closeness of two comparable strength sources prevents the individual measurement of their fluxes. More common for gamma-ray AGN however is the situation where numerous weak sources interfere with the measurement of the flux from a stronger source.

### 2.2.1 Source Confusion and Background Fluctuation Estimates

We will begin our discussion of source confusion by considering a simple model that will indicate the nature of confusion problems, and produce some estimates of the angular resolution needed to avoid these problems. In this model we will consider a simple instrument with a sensitivity  $F_{\text{min}}$  that has a circular field-of-view with solid angle  $\Omega$ . We will assume that the instrument has a flat angular response so that the flux measured is simply the sum of the fluxes from all sources within the field-of-view. We will divide the sources in the sky into two classes; strong sources that have flux greater than the instrument's sensitivity, and weak sources that have flux less than the sensitivity. We will first consider the confusion of strong sources. Then, considering the summed flux of the weak sources as a background, we will estimate the size of statistical fluctuations in this background relative to the sensitivity of the instrument. In section 2.2.2 we consider a more general analysis of source confusion.

**Strong Source Confusion.** If a strong source is centered on the field-of-view then its flux will be confused by the presence of any other strong source in the field-of-view. If we assume that the field-of-view is much smaller than the full sky and that the strong sources are randomly distributed over the sky, then probability of another strong source being in the field-of-view depends only on the mean number of sources per field-of-view. This mean is  $\mu = \Omega N(>F_{\text{min}})$ . If  $\mu$  is small then it in fact gives the probability of confusion by another strong source. Thus these simple considerations impose the resolution requirement

$$\Omega \ll \frac{1}{N(>F_{sen})} \quad (2.2)$$

If for example the instrument is able to detect 100 sources in the full sky and we wish less than 10% to be confused with other strong sources, then the radius of the field-of-view should be less than  $3.6^\circ$ .

**Background Fluctuations.** We will consider the weak sources as contributing to a background that must be subtracted from any flux measurement. This background will vary from field-of-view to field-of-view due to the differing number and flux of the sources that contribute to it. These variations, or background fluctuations, are unmeasurable by the instrument and constitute a source of systematic error in any source flux measurement.

We will suppose that the weak sources are randomly distributed over the sky and that the field-of-view is much smaller than the full sky. The surface density of sources  $N(F)$  gives the distribution of source fluxes.  $N(>F)$  is obtained from  $N(F)$  by integration. We may calculate the mean background flux  $F_B$  measured from

$$F_B = \Omega \int_0^{F_{sen}} N(F) F dF \quad (2.3)$$

while the field-of-view to field-of-view variance of this flux is given by

$$\sigma_B^2 = \Omega \int_0^{F_{sen}} N(F) F^2 dF \quad (2.4)$$

The mean flux depends on the nature of  $N(F)$  for small source fluxes, however the variance depends mainly on  $N(F)$  for sources fluxes within a decade below



the sensitivity of the instrument. In this range we may use a power law approximation to  $N(F)$

$$N(>F) = KF^{-\gamma} \quad (2.5)$$

where as we have mentioned  $\gamma$  is observed to be 1.71. This leads to the estimate

$$\sigma_B^2 = \mu \frac{\gamma}{2-\gamma} F_{sen}^2 \quad (2.6)$$

The  $3\sigma$  flux sensitivity  $F_{sen}$  is three times the statistical error of the flux measurement. Thus since we wish the systematic error due to background fluctuations to be much smaller than the statistical error, we have the following requirement for the angular resolution of the instrument:

$$\Omega \ll \frac{.019}{N(>F_{sen})} \quad (2.7)$$

using 1.71 for  $\gamma$ . As an example, if the instrument is capable of detecting 100 sources in the full sky and we wish the background fluctuation error to be less than half the statistical error, then the instrument's field-of-view must have a radius less than  $0.8^\circ$ . As can be seen, the background fluctuations are an important consideration.

### 2.2.2 The Deflection Distribution

The simple model discussed so far is unrealistic in several respects. Any real instrument will have an angular response more complex than the simple flat response assumed. More important however is the fact the the division between strong and weak sources is completely artificial. We will abandon this division and consider a more rigorous analysis. In this analysis, source confusion and background fluctuations are characterized by a probability distribution called a  $P(D)$  curve, or deflection distribution (Scheuer 1957).  $P(D)$  is the differential probability that the flux measured by the instrument for some randomly selected-direction has a difference or deflection  $D$  from the mean flux (the terminology is from radio astronomy). Since only the mean flux can be subtracted from a source measurement, the  $P(D)$  distribution represents the distribution of noise caused by source confusion.

The relationship between  $P(D)$  and  $N(F)$  was first derived by Scheuer(1957) to characterize the noise in the intensity of low-flux radio objects produced by unresolved sources. Later Scheuer considered the problem for x-ray sources (Scheuer 1974).  $P(D)$  distributions have been employed in studies of fluctuations in the radio and x-ray background ( Condon 1978, Shafer 1983).

**Calculation of  $P(D)$ .** Calculation of  $P(D)$  involves two steps. The first step takes account of the angular response of the instrument. The presence of a source in the instrument's field-of-view induces an increase in its counting rate. This increase will depend on the source's direction  $\vec{\Omega}$  in a manner that will be described by the instrument's response function  $\eta(\vec{\Omega})$ , which we will assume is normalized to one in the direction of peak response. We will mean

by a source's intensity,  $x$ , the flux needed by a source in the direction of the instrument's peak response to produce the same counting rate increase. By the total source intensity  $X$  we will mean the flux needed by a source in the direction of the instrument's peak response to produce the counting rate increase produced by all sources in the instrument's field-of-view. Thus we have

$$X = \sum_k \eta(\vec{\Omega}_k) F_k = \sum_k x_k \quad (2.8)$$

where  $F_k$  is an individual source flux and the sum ranges over all sources for which the instrument has non-zero response.

What we wish to calculate is the distribution of  $D = X - \langle X \rangle$ , the deviation of the intensity from its average over fields of view. This will depend only on the number density  $N'(x)$  defined as the mean number per unit intensity of sources with intensity  $x$ . This is given by

$$N'(x) = \int N(F) \delta(x - \eta(\vec{\Omega})F) dF d^2\Omega \quad (2.9)$$

The second step of the calculation produces  $P(D)$  from  $N'(x)$ . The deflection is the deviation of the total intensity from the mean,  $D = X - \langle X \rangle$ . The technique for determining  $P(D)$  from  $N'(x)$  developed by Scheuer relies on the properties of the Fourier transforms or characteristic functions of probability distributions. It relies on the assumption that sources are randomly distributed in the sky, taking no account of clustering effects. We will simply quote the result (Scheuer, 1974):

$$P(D) = \frac{1}{2\pi} \int CF(\omega) \exp(-i\omega D) d\omega \quad (2.10)$$

where

$$CF(\omega) = \exp \left( \int N'(x) [\exp(i\omega x) - 1 - i\omega x] dx \right) .$$

This result, along with equation (2.9), gives us a general prescription for calculating  $P(D)$  from  $N(F)$ . In general it must be carried out numerically, however for power law  $N(F)$ , portions of the procedure are tractable. We will assume that

$$N(>F) = KF^{-\gamma} \quad 1 < \gamma < 2 . \quad (2.11)$$

In general we must assume that the response function  $\eta(\vec{\Omega})$  can have negative as well as positive values. Gamma-ray measurements require background subtractions, and the background measurements will lead to negative responses. This implies that some sources will have negative intensities, and subtract from rather than add to the total intensity. With this in mind we find

$$N'(x) = \gamma K |x|^{-(\gamma+1)} \begin{cases} \Omega_+ & \text{if } x > 0 \\ \Omega_- & \text{if } x < 0 \end{cases} \quad (2.12)$$

where

$$\Omega_+ = \int_{\eta > 0} [\eta(\vec{\Omega})]^\gamma d^2\Omega$$

and

$$\Omega_- = \int_{\eta < 0} [-\eta(\vec{\Omega})]^\gamma d^2\Omega .$$

Thus for power law  $N(F)$  the effect of the instrument's response is described

by two solid angles  $\Omega_+$  and  $\Omega_-$ , and the intensity distribution is again a power law. This is the same intensity distribution that would be observed by an instrument with a flat response with a solid angle of  $\Omega_+$  for source observation and solid angle  $\Omega_-$  for background observation. From these solid angles we will implicitly define a scaling flux  $F_{scale}$ :

$$(\Omega_+ + \Omega_-)N(>F_{scale}) = 1 \quad . \quad (2.13)$$

On average one source is expected per observation with an intensity greater than  $F_{scale}$  in absolute value. We will also define a response asymmetry ratio

$$R = \frac{\Omega_+ - \Omega_-}{\Omega_+ + \Omega_-} \quad . \quad (2.14)$$

For a single on-source off-source measurement we will have  $R = 0$ , while for a measurement with background subtraction based on a model, we will have  $R = 1$ . With these definitions the deflection distribution may be expressed as

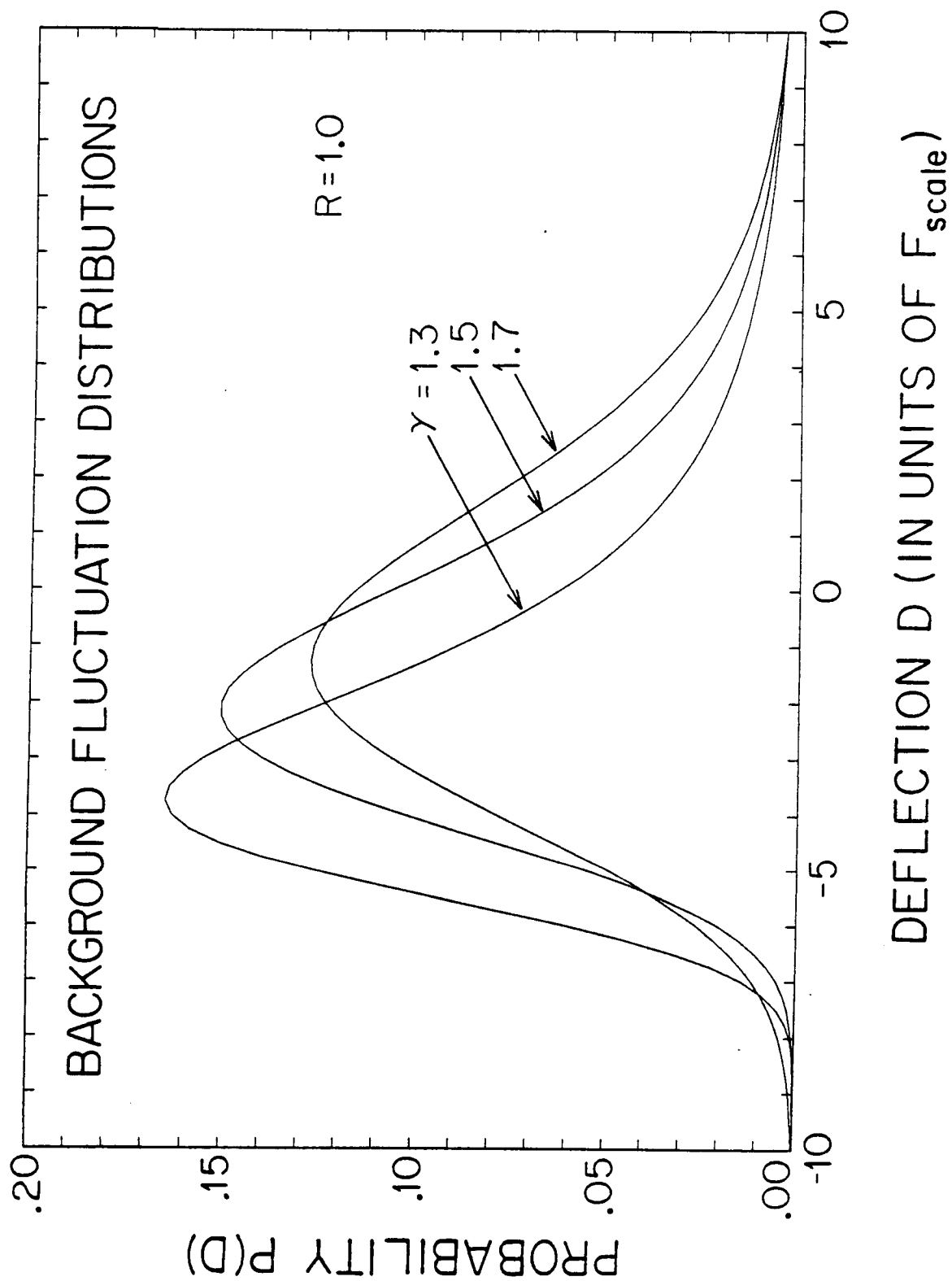
$$P(D) = \frac{1}{F_{scale}} f\left(\frac{D}{F_{scale}}\right) \quad (2.15)$$

where

$$f(u) = \frac{1}{\pi} \int_0^\infty \exp\left[-\cos\left(\frac{\pi}{2}\gamma\right)\Gamma(1-\gamma)\omega^\gamma\right] \cos\left(\omega u - R\sin\left(\frac{\pi}{2}\gamma\right)\Gamma(1-\gamma)\omega^\gamma\right) d\omega \quad .$$

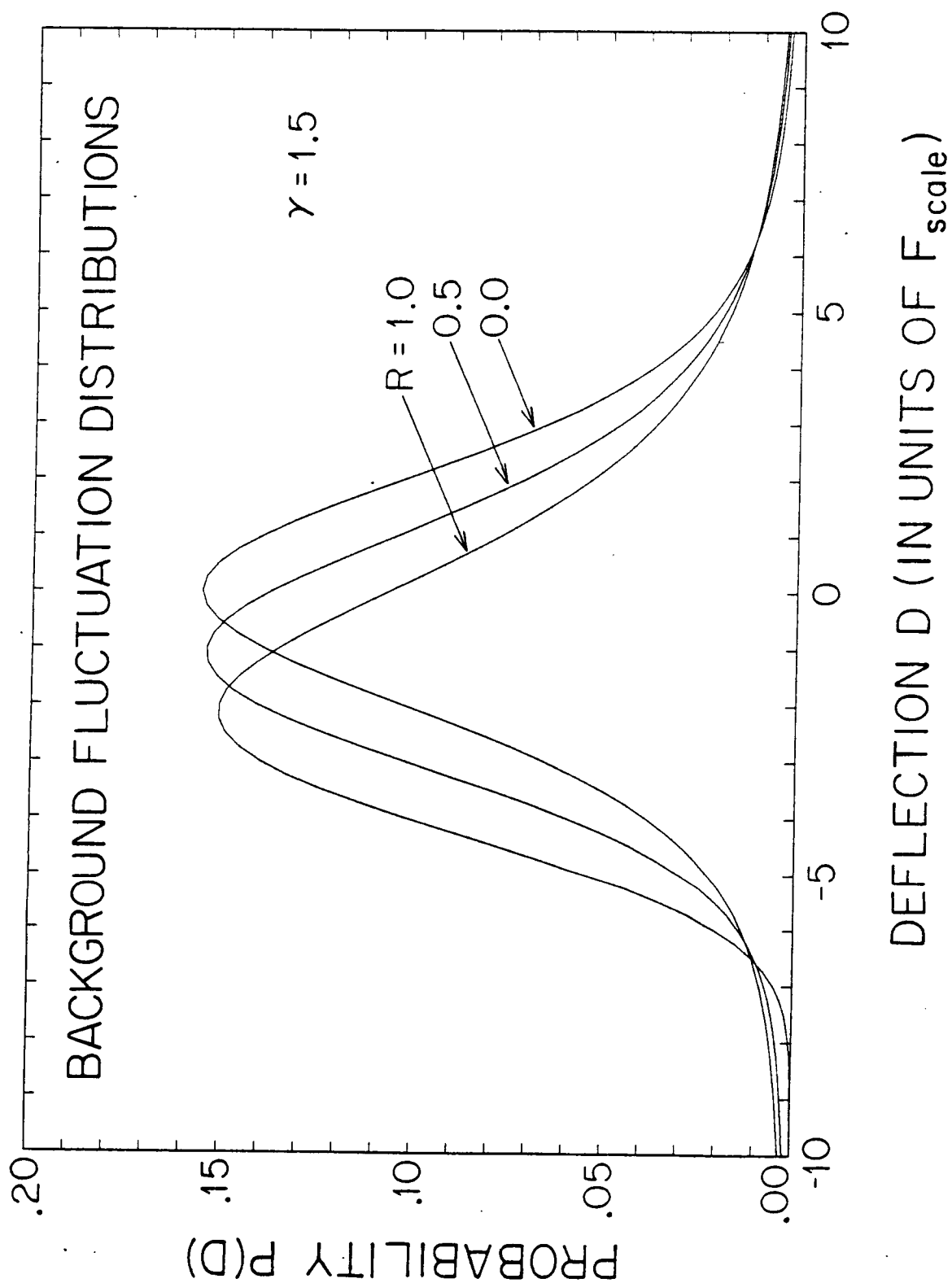
**Description of P(D).** Figures 2.1-2.2 show how this distribution function varies with the  $N(>F)$  slope  $\gamma$  and the asymmetry ratio  $R$ . The  $P(D)$  curves are plotted in units of the scaling flux  $F_{scale}$ . The fact that for a given  $\gamma$  and  $R$  the distributions are a scaling of a universal function is important.

**Figure 2.1.** Deflection distributions for  $R=1$  and several values of  $\gamma$ . These cases correspond to instruments that employ no background measurement, but rather use a model of the background for subtraction purposes.



**Figure 2.2.** Deflection distributions for the Euclidean case  $\gamma=1.5$  and several values of  $R$ .  $R=0$  for instruments that employ a single off-source measurement for background subtraction. If multiply off-source measurements are made for background subtraction, then  $R$  is positive. The limiting case is  $R=1$ , corresponding to a background subtraction based on a model.





We expect on average one source in the field of view with an absolute intensity greater than the scaling flux. The width of the  $P(D)$  distribution scales directly with this flux. The dependence of source confusion effects on the angular resolution of an instrument is expressed through the scaling flux.

Figure 2.1 shows deflection distributions for  $R=1$  corresponding to measurements with no negative response values. This will occur when background subtraction can be based completely on a model of the instrument's background. The peak of each distribution is offset below the mean by an amount that decreases with increasing  $\gamma$ . This offset balances the peak with the high flux tail, which asymptotically approaches  $N'(D)$ . The width of the peak increases with increasing  $\gamma$ , and in fact diverges in the limit  $\gamma \rightarrow 2$ . This is the same divergence that occurs in the estimate of the field-of-view to field-of-view flux variance given by equation (2.4). The divergence is due to the contribution to the variance from fluxes near zero, and can only be cured by altering the lower flux behavior in our model for  $N(F)$ . For an investigation of  $P(D)$  for power law  $N(F)$  with a lower flux cut-off, see Shafer (1983).

Figure 2.2 shows how the distribution varies with the asymmetry ratio  $R$  for power law  $N(F)$  with  $\gamma = 1.5$ , corresponding to a nonevolving population in a Euclidean universe. As  $R$  decreases from 1, a tail appears at fluxes below the mean, while the peak moves closer to the mean. With  $R=0$  the distribution is completely symmetrical. This corresponds to an instrument that uses a single off-source observation for background subtraction. The width and shape of the peak alters very little with  $R$ , with the main effect being in the peak location and the size of the tails.

An important feature of all of these distributions is that the width of the peak is several times  $F_{scale}$ . Thus for example, for the Euclidean case  $\gamma=1.5$  with  $R=1$ , the width of the region that contains 50% of the observations is  $3.65F_{scale}$  while the width of the 90% region is  $10.5F_{scale}$ . The widths of these regions are listed in Table 2.1 for other values of  $\gamma$  and  $R$ .

If we have *a priori* knowledge that there is a strong source in a given direction, the deflection distribution has a simple interpretation. The intensity that will be measured is  $X=F+D$  where  $F$  is the strength of the strong source and  $D$  is the background from all other sources.  $D$  is distributed with the total flux distribution just calculated, so that the width of this distribution translates into an uncertainty in the measurement of the source strength. The uncertainty is due to an uncontrollable systematic variation of the background flux, a basic property of the sky.

For a properly designed instrument, this systematic error should be smaller than the photon statistical error in the measurement of the deflection, and hence smaller than the flux sensitivity of the instrument. The fact that the width of  $P(D)$  is much larger than  $F_{scale}$  therefore implies that a properly designed instrument should have a very small chance of detecting a source in a randomly chosen field-of-view.

**A Model Instrument.** We will consider as an example an instrument with a fairly simple but realistic response function. This will allow us to compare the width of the deflection distribution with the physical parameters of an instrument, and place some requirements on these parameters. We will consider an instrument that is collimated with an "egg-crate" type collimator with FWHM angles  $a$  and  $b$ . To characterize the detector response we will use

50% Width of $P(D)$					
$\gamma$	R = 0.00	R = 0.25	R = 0.50	R = 0.75	R = 1.00
1.3	3.28	3.29	3.31	3.34	3.39
1.4	3.40	3.41	3.43	3.46	3.49
1.5	3.58	3.58	3.59	3.62	3.65
1.6	3.83	3.83	3.84	3.86	3.88
1.7	4.23	4.23	4.24	4.25	4.26
1.8	4.93	4.93	4.93	4.94	4.95

**Table 2.1.** The 50% width in units of  $F_{scale}$  of the background fluctuation distribution  $P(D)$  for various values of  $\gamma$  and  $R$ .

90% Width of $P(D)$					
$\gamma$	R = 0.00	R = 0.25	R = 0.50	R = 0.75	R = 1.00
1.3	12.76	12.65	12.31	11.72	10.98
1.4	11.79	11.71	11.48	11.09	10.63
1.5	11.26	11.21	11.06	10.83	10.55
1.6	11.16	11.13	11.05	10.92	10.76
1.7	11.58	11.57	11.53	11.46	11.38
1.8	12.86	12.86	12.84	12.82	12.78

**Table 2.2.** The 90% width in units of  $F_{scale}$  of the background fluctuation distribution  $P(D)$  for various values of  $\gamma$  and  $R$ .

planar angular coordinates  $x$  and  $y$  oriented along the planes of the collimator. The response can then be described as

$$\eta_{\text{det}}(x, y) = [1 - |\frac{x}{a}|][1 - |\frac{y}{b}|] \quad |x| < a, |y| < b \quad . \quad (2.16)$$

This kind of instrument is generally considered to have a field-of-view of solid angle  $ab$ . For background subtraction the instrument will use observations offset on both sides of the source position. The total response for the source observation, including the background subtraction may be described by

$$\eta(x, y) = \frac{1}{2}[-\eta_{\text{det}}(x+c, y) + 2\eta_{\text{det}}(x, y) - \eta_{\text{det}}(x-c, y)] \quad (2.17)$$

where  $c$  is the offset for the background pointings. We will assume that  $c = 1.25a$  so that there is substantial overlap of the source and background fields of view, but with no overlap in the center of the source field. This overlap reduces the susceptibility of observations of known sources to background fluctuations. A contour plot of this response function is shown in figure 2.3.

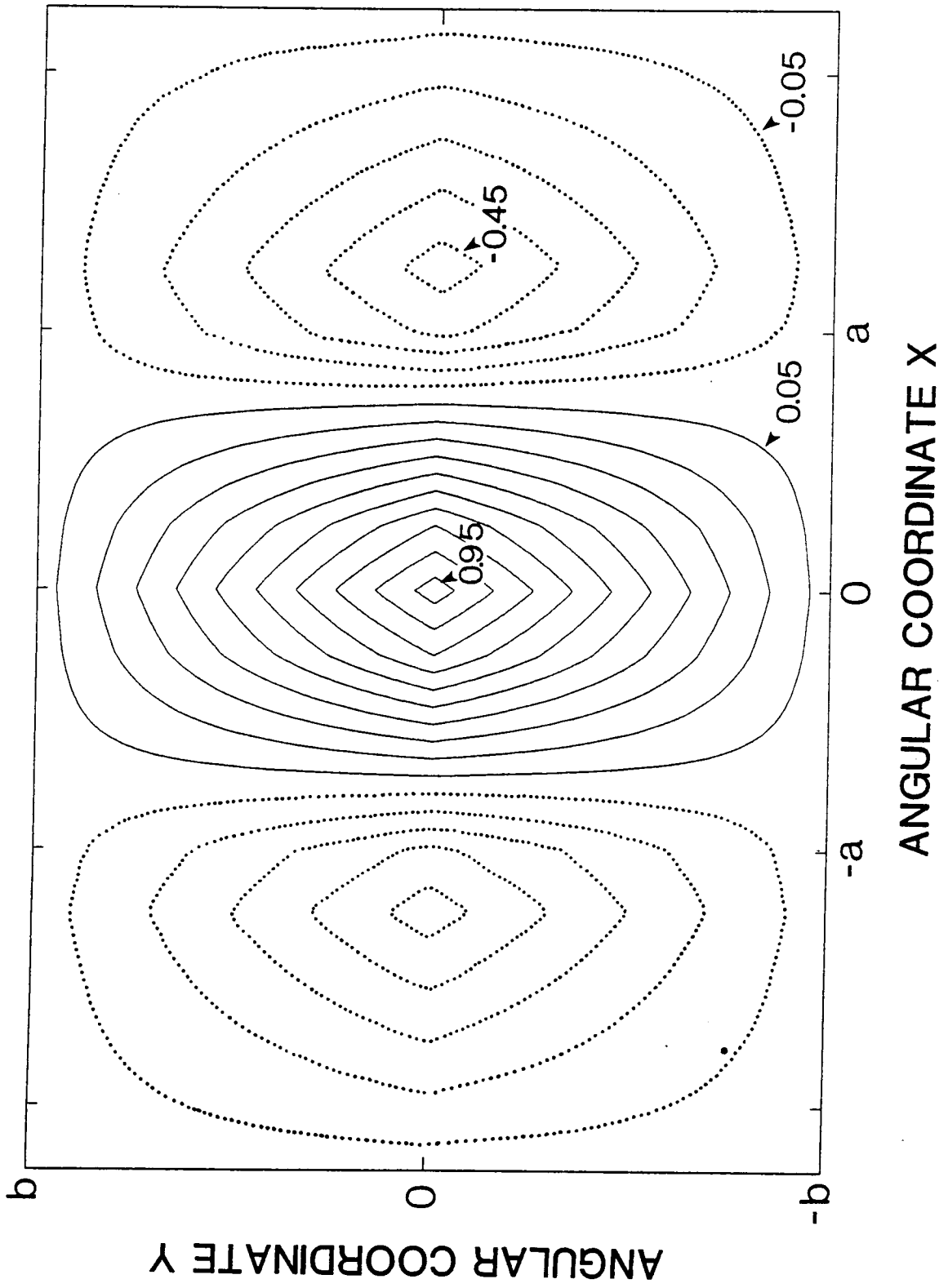
For  $\gamma = 1.71$  we find that

$$\Omega_+ + \Omega_- = .743 \cdot ab, \quad R = .24122 \quad . \quad (2.18)$$

This total solid angle increases by 20% for non-overlapping background fields-of-view, while the asymmetry ratio is independent of the offset  $c$ .

**Resolution Requirements.** For this model instrument we can now consider the requirements for the angular resolution. We wish to insure that the width of the fluctuation distribution  $P(D)$  is small compared to the statistical error of a flux measurement. To do this we will compute an effective

**Figure 2.3.** Contour plot of the instrument response function for an instrument with "egg-crate" type collimation using two offset pointings for background subtraction. The peak response at the center is normalized to one. Contour steps are in intervals of 0.1, with positive contours being solid and negative contours dotted. Negative responses occur in regions used predominantly for background subtraction.



width  $\sigma_{fluct}$  for the  $P(D)$  curve. For the parameters  $\gamma=1.71$  and  $R=2.4122$  the 90% width of the  $P(D)$  curve is the same as the 90% width of a Gaussian with sigma  $\sigma_{fluct} = 3.54F_{sen}$ . We wish this to be much less than the statistical error  $\sigma_{stat}$ , which we will take to be a third of the  $3\sigma$  flux sensitivity of the instrument  $F_{sen}$ . We find for this example that

$$ab = \frac{.024}{N(>F_{sen})} \left( \frac{\sigma_{fluct}}{\sigma_{stat}} \right)^{\gamma} \quad (2.19)$$

requiring  $1.7^\circ$  by  $1.7^\circ$  collimation if we wish to observe 100 sources and reduce the fluctuation error to half the statistical error.

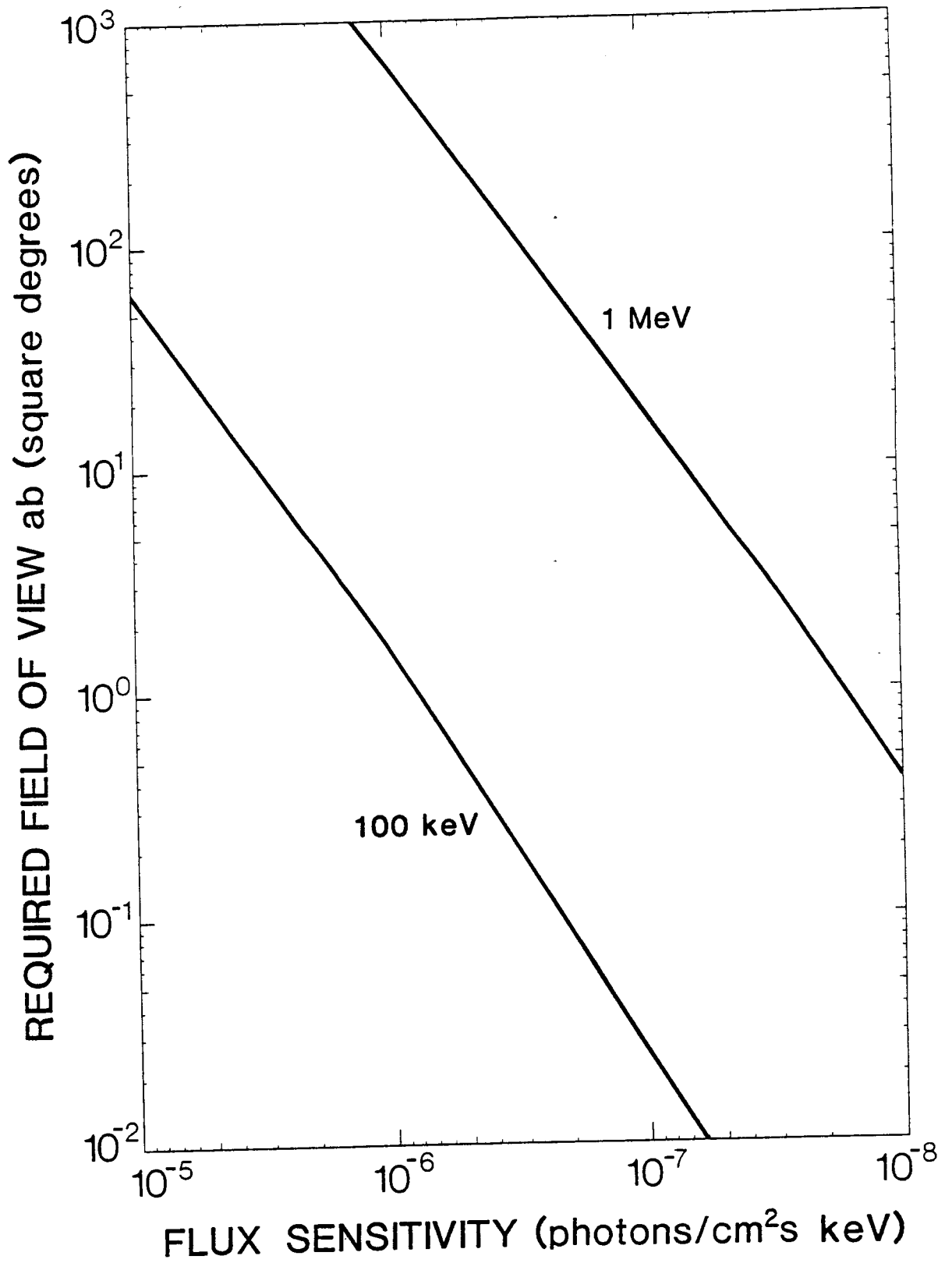
Figure 2.4 shows the solid angle of the field-of-view  $ab$  that is required to reduce the fluctuation error to half of the statistical error versus an instrument's  $3\sigma$  sensitivity. This is shown for both 100 keV and 1 MeV and is based on the  $N_{AGN}( >F)$  estimated presented in Chapter 1. The curves predict that the OSSE instrument which is being developed for the GRO satellite and uses an egg-crate type collimator with a 35 square degree field-of-view does not have sufficient resolution to achieve its 10-day observational sensitivities (Kurfess et al. 1983) of  $3.5 \cdot 10^{-7}$  photons/cm<sup>2</sup>·s·keV at 100 keV and  $5 \cdot 10^{-8}$  photons/cm<sup>2</sup>·s·keV at 1 MeV.

### 2.2.3 Source Detection

Survey instruments which actively seek out previously unknown sources face the problem of separating real sources from statistical fluctuations and background fluctuations. If the background fluctuations are small compared to the photon statistical flux uncertainty, then the detection problem reduces



**Figure 2.4.** Degree of collimation required to achieve a given sensitivity. The curves give the maximum allowed product  $ab$  of the FWHM angles  $a$  and  $b$  of an egg-crate collimator versus the  $3\sigma$  instrument sensitivity, at 100 keV and 1 MeV, and are based on the estimates of  $N(>F)$  presented in chapter I. The curves represent the solid angle for which the effective width  $\sigma_{fluct}$  of the background distribution is half of the statistical error  $\sigma_{stat}$  of the flux measurement.



to the choice of a flux threshold. If the flux measured from a given region of the sky is higher than the threshold, the field is assumed to contain a source, otherwise it is not. With a proper choice of threshold, the risk of claiming that a source exists when it does not is small, while sources actually above the threshold have a reasonable chance of being successfully detected.

At source strengths where background fluctuations are important this simple picture of the detection problem breaks down. The simple assumption that the field-of-view being observed contains a uniform background with or without a single source is no longer an appropriate approximation. In reality the field-of-view contains a large number of sources, and yields to no simple description. At these low-flux levels it is unclear whether we are observing the effect of a single source, several sources, or the collective effect of a large number of sources. If an instrument measures a flux from a given direction greatly above the average, we expect that in this direction there is a source with strength near this excess. If however the excess is near the width of the background fluctuation distribution we can no longer make this interpretation. In this section we will examine the cases for which we may still assume that the excess is due to a single source.

The natural candidate for this single source is that source which makes the largest contribution to the total flux. Experimentally we would like to know after we measure a deflection  $D$  how likely it is that this largest contribution  $x_{\max}$  accounts for the excess. If this is very likely then the interpretation is straightforward, otherwise we must consider multiple source models for the excess.

What we shall determine is the conditional probability  $P(x_{\max}|D)$  of having a largest intensity contribution  $x_{\max}$  given that the total deflection measured is  $D$ . Given this distribution we can evaluate our confidence that there is a single source that accounts for the excess. To employ the machinery used to calculate the background fluctuation distribution, we will make use of the fact that

$$P(x_{\max}|D) = P(D|x_{\max}) \frac{P(x_{\max})}{P(D)} . \quad (2.20)$$

The first term in this expression is the distribution of  $D$  given the *a priori* knowledge that the largest source has intensity  $x_{\max}$ . There is a related distribution  $P(D'|x < x_{\max})$  that depends on the weaker assumption that all source intensities  $x$  are less than  $x_{\max}$ . This may be calculated using equations (2.10) by substituting for the intensity count distribution

$$N_{x_{\max}}'(x) = \begin{cases} N'(x) & x < x_{\max} \\ 0 & x > x_{\max} \end{cases} \quad (2.21)$$

where the restriction of  $N'(x)$  to  $x < x_{\max}$  accounts for the knowledge that all sources have intensities below  $x_{\max}$ . The result of this calculation  $P(D'|x < x_{\max})$  is the distribution of the difference  $D' = X - \langle X \rangle'$  between total intensity and the average total intensity given that all source intensities are less than  $x_{\max}$ . This difference is related to the deflection  $D$  by

$$D' = D + \langle X \rangle - \langle X \rangle' = D + \int_{x_{\max}}^{\infty} N'(x) x dx . \quad (2.22)$$

The presence of the source intensity  $x_{\max}$  increases the total deflection  $D$  by

$x_{\max}$  so that

$$P(D|x_{\max}) = P(D - \langle X \rangle + \langle X \rangle' - x_{\max} | x < x_{\max}) . \quad (2.23)$$

The remaining term  $P(x_{\max})$  may be simply evaluated. If there are no sources with intensity greater than some limit  $x_{\lim}$  then it follows that  $x_{\max}$  is less than  $x_{\lim}$ . Thus we can equate the probability that there are no sources with intensity greater than  $x_{\lim}$  with the probability that  $x_{\max}$  is less than  $x_{\lim}$  :

$$\exp(-N'(>x_{\lim})) = \int_{-\infty}^{x_{\lim}} P(x_{\max}) dx_{\max} \quad (2.24)$$

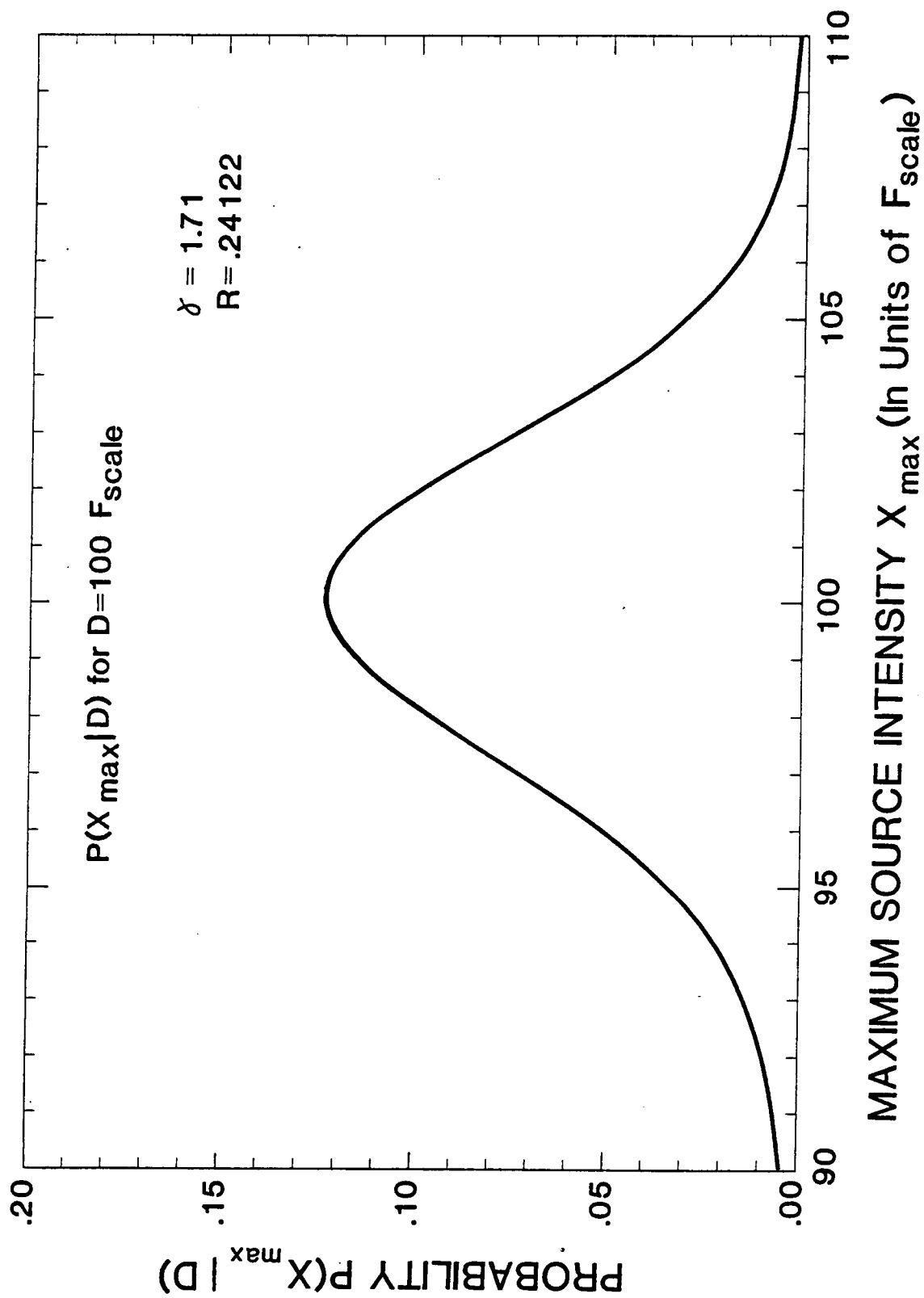
so that

$$P(x_{\max}) = \exp(-N'(>x_{\max})) N'(x_{\max}) . \quad (2.25)$$

Figures 2.5 and 2.6 show calculated distributions of the largest source contribution  $x_{\max}$  for various deflections. For the calculation we have assumed a  $\gamma = 1.71$  power law for the source number distribution, and used the response function of section 2.2.2. Fluxes in the figure are given in units of the  $F_{scale}$ . Each curve is labeled with the value of  $D$  in units of  $F_{scale}$ .

For large flux excesses we expect that the largest contribution is approximately given by  $D$ , with the distribution about this value being close to the fluctuation distribution mirrored about zero. This simply says that if we are sure that the excess is a good measure of the strength of a single source, then the errors on this measurement are given by the background fluctuation distribution. Figure 2.5 shows the distribution of  $x_{\max}$  for  $D = 100F_{scale}$ , verifying this expectation.

**Figure 2.5.** The probability  $P(x_{\max}|D)$  of a field-of-view with deflection  $D=100F_{scale}$  has a maximum source intensity  $x_{\max}$ . For large deflections such as this, the distribution of  $x_{\max}$  about  $D$  is the mirror image of the  $P(D)$  distribution. The curve is calculated assuming  $\gamma=1.71$ .



**Figure 2.6.** The probability  $P(x_{\max}|D)$  of a field-of-view having a maximum source intensity  $x_{\max}$  given various values of  $D$ . Each curve is labeled with the value  $D$  in units of  $F_{\text{scale}}$ .



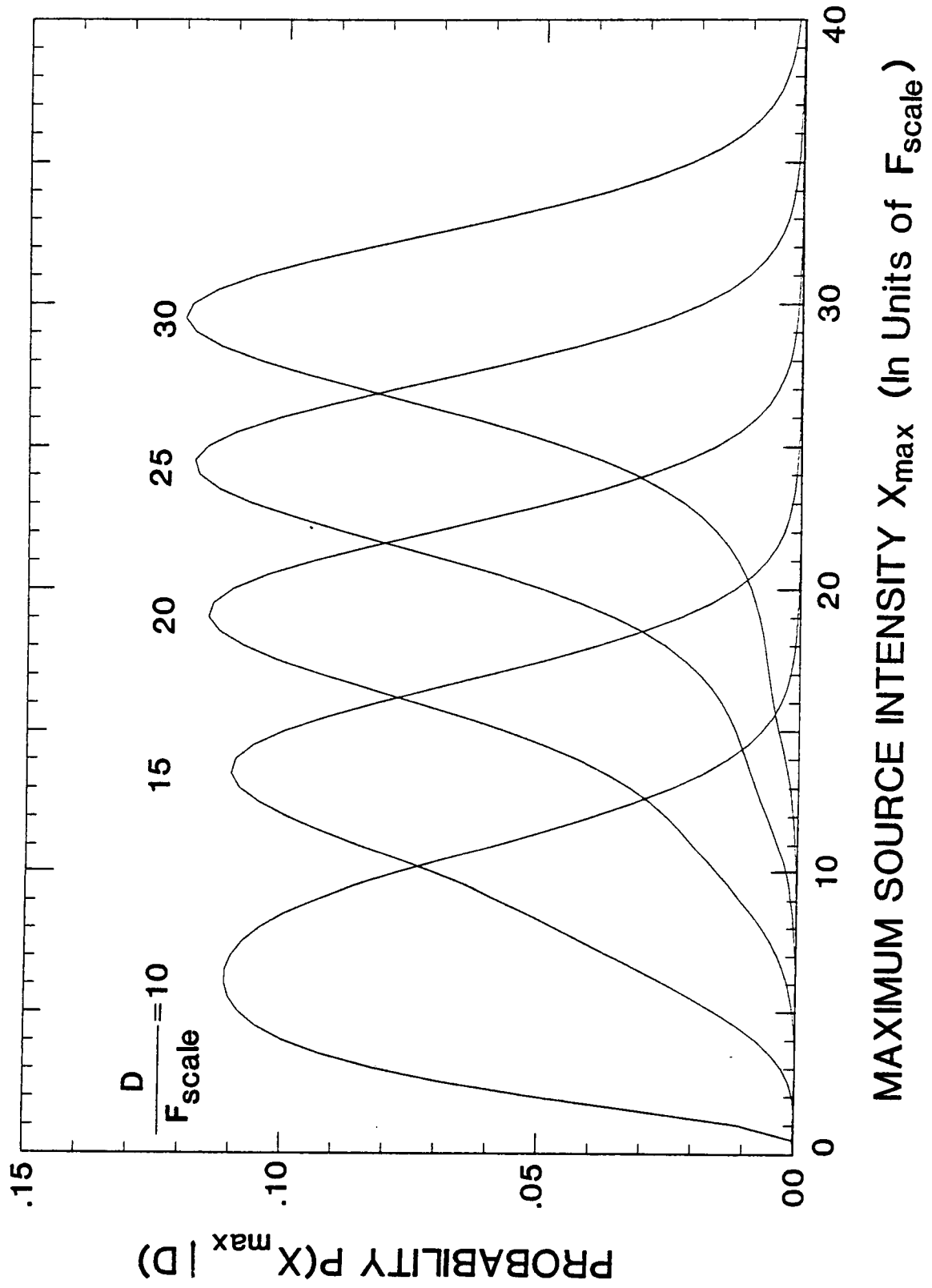
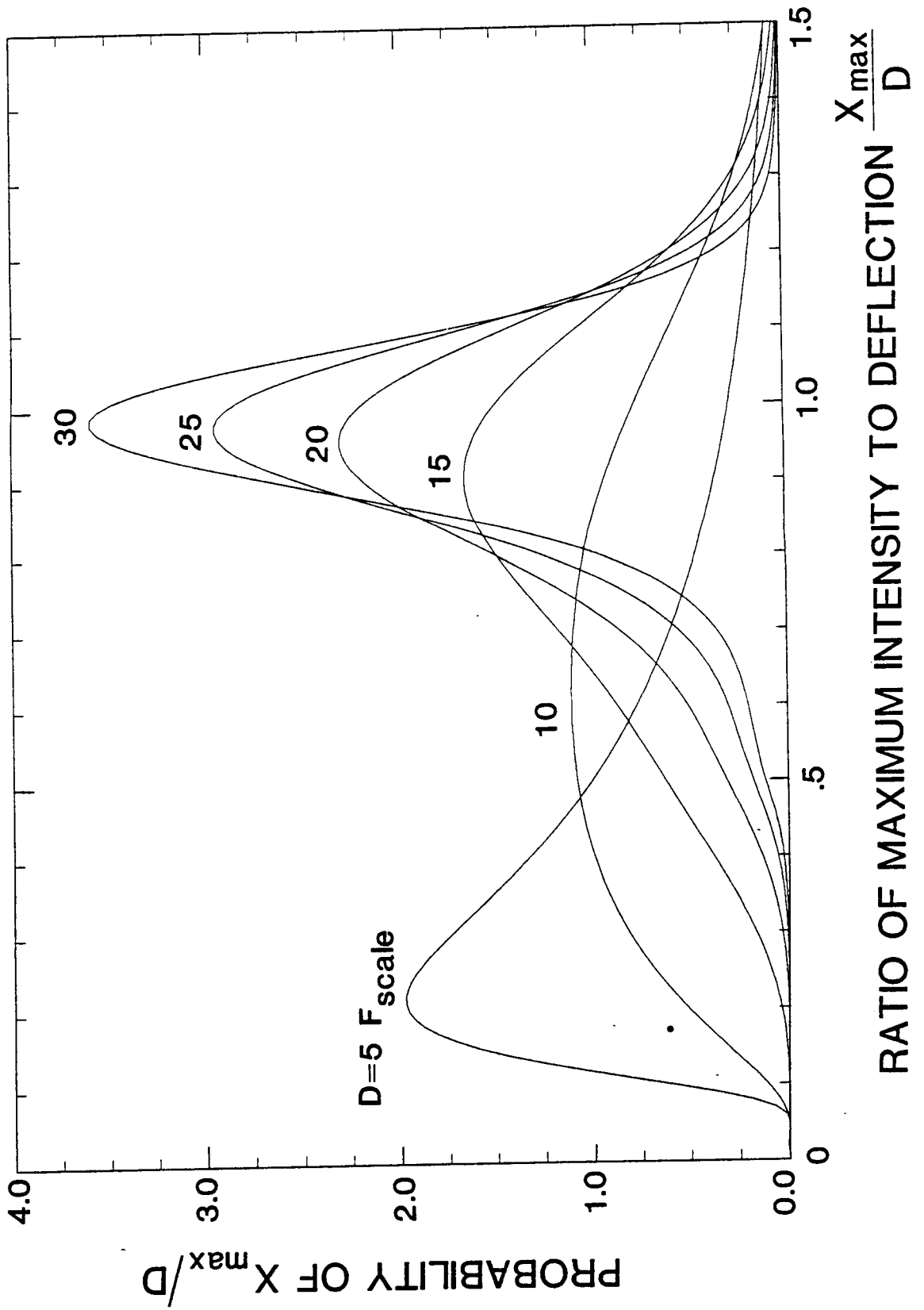


Figure 2.6 shows  $P(x_{\max}|D)$  distributions for small deflections  $D$ . At  $D = 30F_{scale}$  the distribution is still approximately given by the reflection of the  $P(D)$  distribution, but two modifications can be noted. First the peak is shifted to below  $x_{\max} = 30F_{scale}$ . This is due to the fact that it is more likely to encounter a weaker source with a compensating stronger background fluctuation, than a stronger source and weaker background fluctuation. For decreasing deflection  $D$  the relative magnitude of this bias increases. The second modification is in the low source flux tail of the distribution, which is enhanced in the region  $x_{\max} > \frac{1}{2}D$ , and diminished below  $\frac{1}{2}D$ . In this tail region the deflection is primarily due to two sources of comparable intensity. As  $D$  decreases from  $30F_{scale}$  this tail region grows, and near  $D = 10F_{scale}$  dominates the single source peak.

Figure 2.7 examines this behavior in more detail. The plot shows the probability of observing the ratio of  $x_{\max}$  to  $D$ . Each curve is labeled with the value of  $D$  in units of  $F_{scale}$ . The growth with decreasing  $D$  of the two-source region of the distribution is clearly seen. In addition it is apparent that for deflections  $D$  below  $15F_{scale}$ , the pile-up of three or more sources becomes an ever more likely explanation for the deflection.

In the presence of this source confusion we will use the following formulation for the detection problem. We will choose a limit  $x_{lim}$ , and call fields-of-view with maximum source intensity  $x_{\max}$  greater than this limit "source fields", while fields-of-view with maximum source intensity less than this limit will be called "background fields." How well can we distinguish source fields from background fields based solely on the measured deflection  $D$  ?

**Figure 2.7.** The probability of a field-of-view with deflection  $D$  having a given ratio of the maximum source intensity  $x_{\max}$  to  $D$ . Each curve is labeled with the deflection  $D$  in units of  $F_{scale}$ . For small deflections the peak moves away from  $x_{\max}=D$  to lower values, indicating that the deflection is most likely due to several sources.



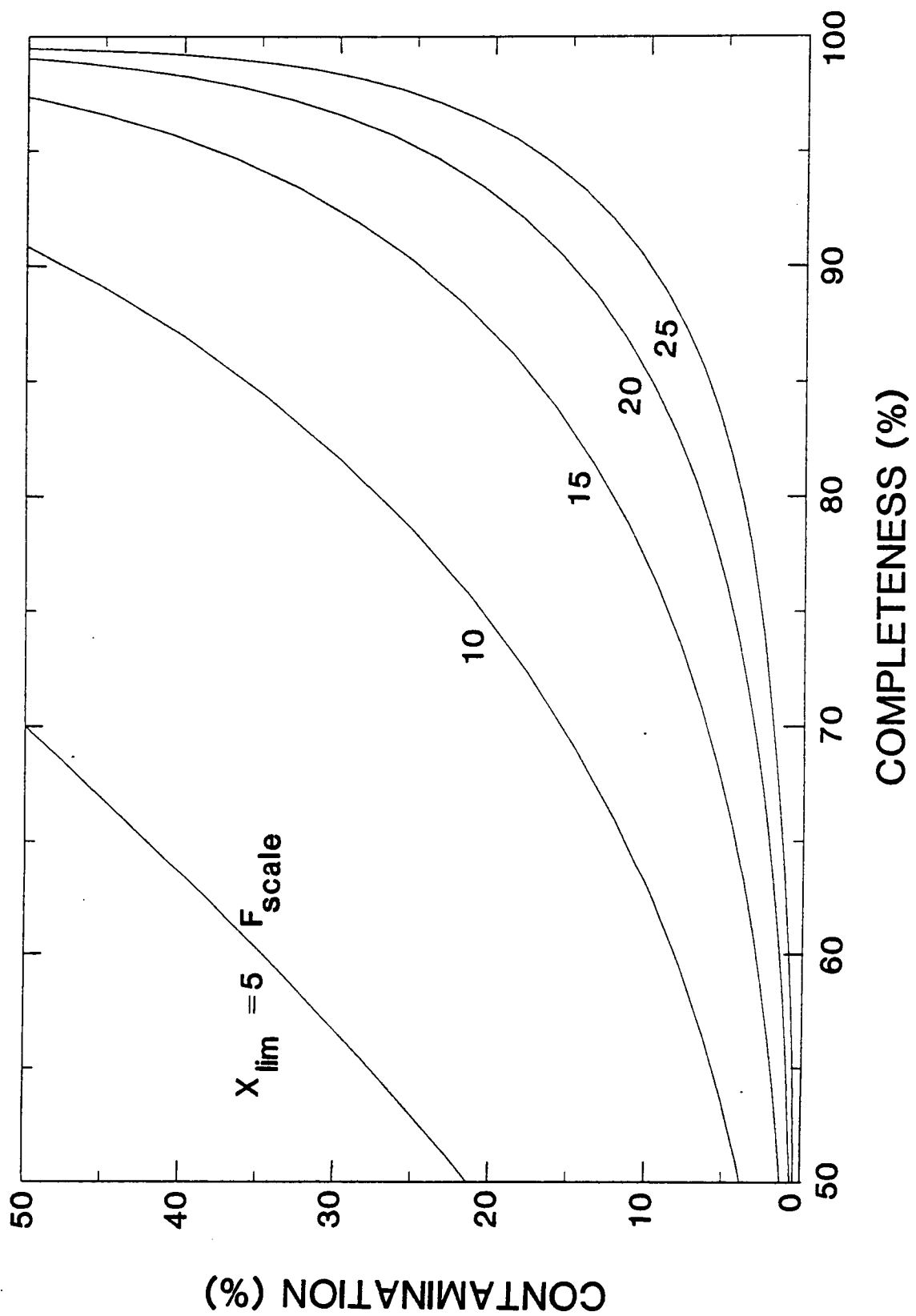
We will choose a deflection threshold  $D_{th}$  and decide that a field-of-view is a background field if the deflection is below this value, and a source field if it is above. There are two important criteria for evaluating this source field detection process. First we must consider the completeness, or the fraction of source fields that are identified as source fields (this is also called the true positive fraction). Second we must consider the degree of contamination, the fraction of claimed source fields that are in fact background fields.

The completeness and the contamination depend on the choice of the deflection threshold  $D$ . By changing this threshold we may increase the completeness at the cost of greater contamination, or we may decrease the contamination at the cost of lower completeness. Figure 2.8 shows the relationship between the contamination and the completeness for the instrument model of section 2.2.2, for the extreme case of no statistical uncertainty on the measured deflection  $D$ . The curves are labeled with the value of  $x_{lim}$  in units of  $F_{scale}$ , and can be derived from  $P(D)$  and  $P(D'|x < x_{lim})$ .

As can be seen, our ability to separate source fields from background fields is poor for low source intensity limits, but improves at higher limits. For limits above  $25F_{scale}$  it is possible to have 90% completeness with less than 10% contamination. Since the statistical error associated with the determination of  $D$  in this analysis can only degrade the detection performance, we will consider as a minimal requirement for a survey instrument that  $x_{lim} < 25F_{scale}$ . This requires of the angular resolution that

$$ab < \frac{.0055}{N(>x_{lim})} \quad (2.25)$$

**Figure 2.8.** Survey completeness versus contamination. For each field-of-view in the survey a decision is made as to whether it contains a source with intensity  $x_{\text{lim}}$  or greater. This decision is based on comparison of the measured deflection to a threshold value  $D_{th}$ . The completeness is the percentage of fields-of-view that actually contain a source of intensity greater than  $x_{\text{lim}}$  that are correctly identified. The contamination is the percentage of fields-of-view identified as containing a source with intensity greater than  $x_{\text{lim}}$  that in fact do not. Each curve is labeled with the value of  $x_{\text{lim}}$  used, while different points along a curve correspond to different values of  $D_{th}$ .



or for the detection of 100 sources in the full sky less than a  $1.5^\circ$  by  $1.5^\circ$  field-of-view.

## 2.3 Conclusions

In this chapter we have considered several issues that impact the angular resolution requirements of a gamma-ray instrument. We have considered the problem of finding counterparts for gamma-ray sources, and considered the source confusion effects on source flux estimation and source detection. For each of these problems it is the population of sources with fluxes within a decade below the instrument's sensitivity that must be considered, and it is the density of such sources that drive the resolution requirements.

There are several problems that we have not considered, but which are relevant to instrumental angular resolution. First the measurements we have considered consist of one or several fixed pointings of a relatively simple instrument. We have not considered instruments that scan across the sky. We can apply our results to the instantaneous response of such an instrument, however the process of searching a continuous set of deflections for peaks has not been considered. This process will produce a bias toward larger background fluctuations. However, we expect only minor changes in the instrument's resolution requirements. We also have not considered more complex instruments that can measure more than one property of a given field-of-view. Such an instrument has several response functions, and our results may be applied to them individually. Correlations between the background fluctuations in each deflection will exist if there is any overlap in the responses. A major concern of this thesis is the development of imaging



instruments which have a large number of responses with little overlap.

A second problem that we have not considered is the temporal fluctuations of the discrete source background. This is a concern in the observation of the temporal behavior of a fixed source. Since the underlying population of unresolved AGN are known to be variable, their summed flux will also be variable. We can make a simple estimate of the scale of this variability. We will characterize the variability on each source by a temporal variance  $\sigma_F^2$ , and suppose that the average of  $\sigma_F^2$  over all sources with flux  $F$  is proportional to  $F^2$ . For a Euclidean non-evolving universe this proportionality is strictly true, however for an evolving or non-Euclidean universe correlations between source luminosity and variability may introduce a slow dependence of the proportionality constant on the source flux. In this approximation the expected temporal variance of the total flux in a field is proportional to the expected sum of  $F^2$  for sources in the field-of-view. Inspection of equation 2.4 then reveals that this average temporal variance is proportional to the field-of-view to field-of-view variance in the background flux. Thus temporal fluctuations scale with the field-of-view to field-of-view variations of the background.

These above considerations therefore do not alter our basic conclusion. Gamma-ray instruments must be capable of angularly resolving sources with fluxes a decade or so below their sensitivity, even though these sources cannot be detected. Major advances in gamma-ray astronomy will require instruments sensitive to perhaps 100 sources. For such instruments to realize their full potential, they must have degree or sub degree angular resolution.

### 3. Gamma-Ray Imaging

Significant improvements of our understanding of active galaxies and the gamma-ray background will require gamma-ray instruments capable of observing perhaps 100 sources. In chapter 1 we estimated the sensitivity required of such instruments, while in chapter 2 we discussed their angular resolution requirements. In this chapter we will argue that achievement of these sensitivities and angular resolution will require the development of new observational techniques. We propose that future gamma-ray instruments will require imaging capabilities, being able to determine numerous source fluxes from subregions of their total field-of-view. After considering the advantages of such imaging capabilities, we will consider one imaging technique, coded aperture imaging, in detail.

One of the major implications of the angular resolution requirements established in chapter 2 is that a properly designed non-imaging instrument should detect less than one source per 180 randomly selected observations. This requirement has a major impact on the design of survey instruments. In a simple single response instrument the angular resolution and the size of the field-of-view are closely related. The number of observations needed for such an instrument to complete a full sky survey with sub-degree angular resolution require a prohibitive amount of time. In an imaging instrument the field-of-view is divided into subregions with a flux measured for each region. The angular resolution of an imaging instrument is determined by the size of

these subregions and not by the size of the field-of-view. In this chapter we will show that the flux sensitivity for each subregion can be nearly the same as the instrument would have for the whole field-of-view if it were deprived of its imaging capabilities. Thus the addition of the imaging capability has little effect on the instrument's flux sensitivity.

The time for an instrument to complete a full sky survey to a given sensitivity is proportional to the time needed to reach that sensitivity for a single observation, and is inversely proportional to the solid angle of the field-of-view. For a simple single response instrument with sub-degree angular resolution, more than ten thousand observations are needed to cover the sky. For instruments with such fine collimation, the instrumental background and hence the time needed to reach a given sensitivity is insensitive to even large changes in the solid angle of the field-of-view. Thus increasing the size of the field-of-view, and adding imaging capabilities to recover the angular resolution, will greatly decrease the time needed to complete a sky survey.

The desirability of imaging capabilities is not restricted to survey instruments. Any gamma-ray observation is conducted in the presence of a background of events from a variety of sources. These include environmental photons that leak through shielding elements, photons from the nuclear activation of the gamma-ray detector and nearby materials by cosmic-ray particles, and neutron interactions. For the sensitivities we are discussing the event rate due to this background can be a thousand or more times the rate induced by a source.

At these low signal-to-noise ratios the background subtraction must be done with great care. In particular the conventional on-source off-source

technique is inappropriate. In this technique the rate is measured with the instrument pointing at a presumed "source field" and then with the instrument pointing at a presumed "background field," with the subtracted rates interpreted as due to a source. Even small changes of the instrumental background with orientation or time can cause large systematic errors in the source measurement. Such variations may be due to the directional dependence of environmental backgrounds or relative changes in the orientation of masses near the photon detector.

In this chapter we will discuss an imaging technique that will allow a background subtraction that is less prone to systematic error. These subtractions happens at two levels: 1) an instantaneous comparison of different portions of the gamma-ray detector, and 2) shortly separated on-source off-source measurements for each portion of the detector with no need for change in the orientation of the instrument or change in the distribution of mass near the instrument. The multiple levels of subtraction possible with an imaging instrument insure flux measurements less prone to systematic biases.

No guarantees exist that unforeseen systematic effects will not occur. With a simple single response instrument, there are few ways to test for such effects. In any observation an imaging instrument will measure many more fluxes than the number of sources expected to be detected. The source fluxes are then highly over-determined and the remaining fluxes provide a consistency check to detect the presence of systematic effects. Hypotheses about the instrument background as well as the source configuration may be tested using the predicted effects on the numerous fluxes measured.

### 3.1 Coded Aperture Imaging

Coded aperture imaging is a technique for imaging sources of photons in the x-ray and gamma-ray energy region where mirrors and lenses are not available. The technique employs a mask composed of transparent and opaque regions which is interposed between the photon sources and a position-sensitive detector. The flux from any given source is spatially modulated by the mask forming a shadow pattern on the detector. If the shadow produced by a source in any given direction in the field-of-view is unique, we can determine the source distribution from the spatial modulation of the flux incident on the detector. For an introduction to the literature on coded aperture imaging the reader is directed to the review by Skinner (1984) and references therein, and also the books by Barrett and Swindell (1981) covering work with coded aperture imaging in nuclear medicine.

In an astronomical observation the detector is subjected to a large background from external photons and internal interactions. This background is likely to vary with position, making the detection of the source flux modulation difficult and prone to error (e.g. McConnel et al. 1982). This problem is solved by employing an anti-mask, a mask based on the same pattern but with transparent and opaque regions interchanged. Substitution of the mask with the anti-mask leaves the background unaltered but reverses the source flux modulation. The difference between the fluxes detected with the mask and anti-mask is thus independent of the position-dependent background.

Considerable work has been done investigating masks suitable for imaging. Dicke (1968) initially proposed masks with a random array of holes. More recent work has focused on masks which have better imaging properties,

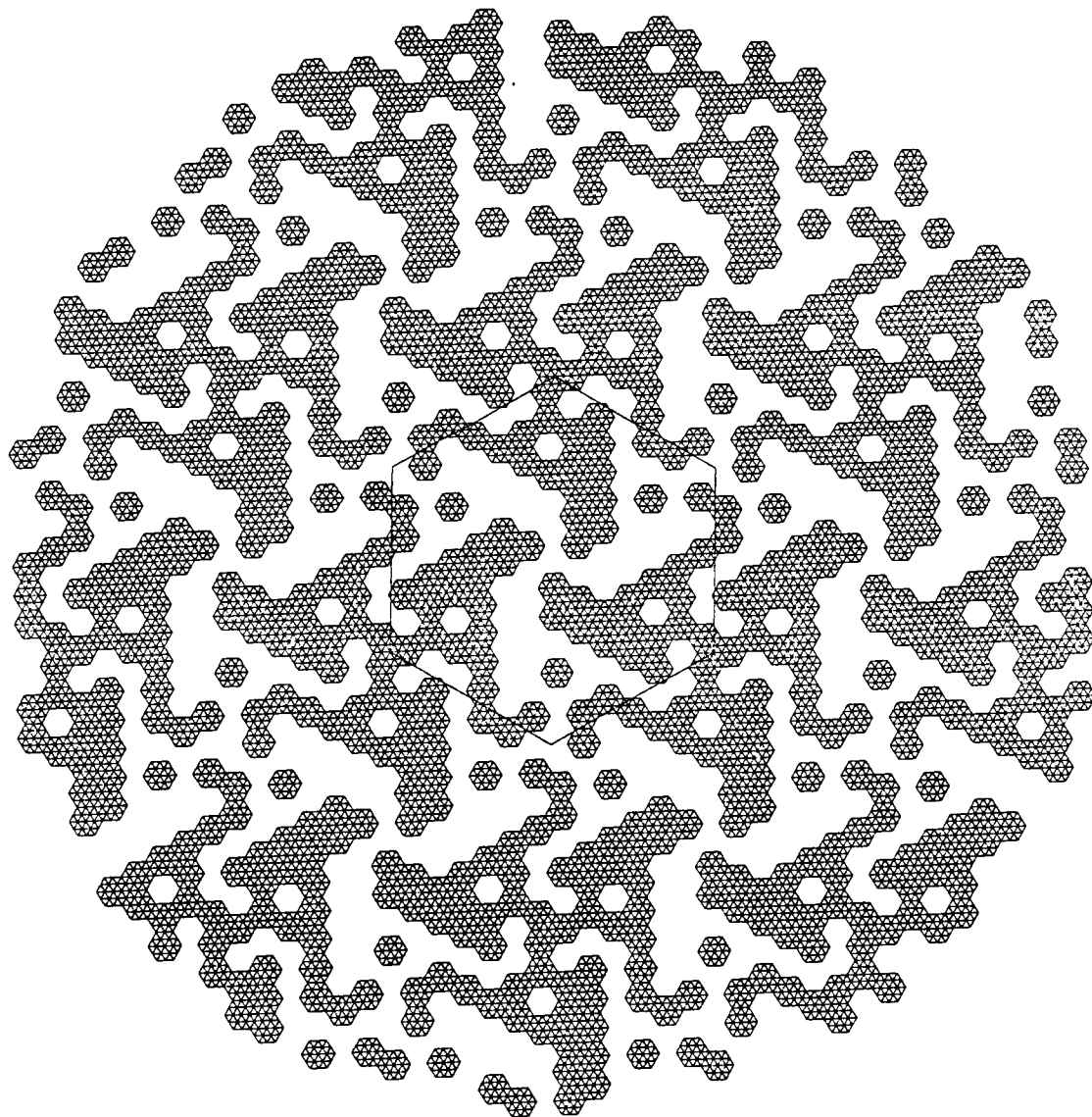
commonly known as Uniformly Redundant Arrays (Gunson and Polychronopoulos 1976, Fenimore and Cannon 1978).

In this chapter we will present an analysis of the coded aperture imaging of point sources of gamma-rays with instruments employing masks based on Hexagonal Uniformly Redundant Arrays (HURAs) (Cook et al. 1984, Finger and Prince 1985). These masks have many properties that make them attractive for gamma-ray astronomy. An example of such a mask is shown in Figure 3.1. The general prescription for the construction of such a mask is given in Finger and Prince (1985). These masks consist of a plane array of hexagonal cells with each cell being fully opaque or transparent. The cells are opaque or transparent in accordance with the periodic pattern of an HURA. In practice this pattern is finite; we will however ignore the effects due to the edge of the mask by considering an HURA pattern and its repetitions. Within a single repetition of the HURA pattern there are  $v$  cells of which  $\frac{1}{2}(v+1)$  are transparent and  $\frac{1}{2}(v-1)$  are opaque. For the mask in Figure 4.1,  $v=127$ . This half-open, half-closed property provides the optimal flux sensitivity in a high background environment.

The uniform redundancy of the pattern is a translational property which guarantees that the shadow pattern cast by a given source is unique to that source. This property involves the frequency with which pairs of opaque cells with a given separation occur. If  $\Delta\vec{r}$  is a possible spacing between cell centers and does not translate the mask pattern onto itself, then exactly  $(v-3)/4$  of the  $(v-1)/2$  opaque cells from a single repetition of the pattern are spaced by  $\Delta\vec{r}$  from other opaque cells in the mask. This property is important in obtaining ghost-free images.

**Figure 3.1.** A hexagonal uniformly redundant array of order 127. The basic pattern which is repeated is composed of 127 hexagonal cells. The pattern is almost anti-symmetric upon rotation by  $60^\circ$ , with opaque cell interchanging with transparent cell. This allows mask rotation to be used for accurate background subtraction.

ORIGINAL PAGE IS  
OF POOR QUALITY





An HURA pattern is also anti-symmetric upon rotation by  $60^\circ$ , with transparent and opaque cells interchanging, except for the central cell and its repetitions. This allows a simple solution to the mechanical problem of interchanging mask and anti-mask. A key feature in the approach described in this chapter is the introduction of time modulation into the image coding process through the continuous rotation of the HURA mask. This continuous rotation allows extension of the field-of-view by repetition of the basic HURA pattern. The rotational motion eliminates the ambiguity that would normally result with a stationary mask having multiple repetitions of a basic pattern. This enlargement of the field of view and corresponding increase in the number of pixels in the image can be accomplished with no increase in the size or resolution of the detector.

We will begin our analysis of the decoding of the flux modulation produced by an HURA mask by considering observations that involve only two mask orientations, corresponding to a mask and anti-mask. In section 3.2 we develop a model of the modulation of the source flux produced by the coded mask. This model includes the effects of the finite resolution of the detector.

In section 3.3 we consider the estimation of source fluxes and directions from fluxes detected with the mask and anti-mask observations. From a maximum likelihood analysis emerges a continuous image, the correlation image. This image is a continuous reconstruction of the source distribution, formed by correlating the detected flux modulation with the modulation expected from a point source in a given direction. It contains all the information in the data needed to determine source fluxes and directions. Except for the mask-anti-mask subtraction involved, this image is similar to the

0-2

continuous image considered by Fenimore (1980).

In section 3.4 we consider the statistical errors associated with source fluxes and locations determined from the correlation image. This analysis differs from that presented by Fenimore (1978) or Caroli et al. (1984) in that the image considered is continuous rather than discrete, and the effects of the finite position resolution of the detector have been taken into account. Our results are then applied to the problem of choosing the mask cell size to minimize source flux errors or source location errors.

In section 3.5 we extend the analysis to include continuous rotation of the mask. We show how the rotation eliminates the ambiguity due to the periodic nature of the mask. This allows the observation of sources over a much larger field-of-view, with no change in the requirements on the detector. We present images that demonstrate the effectiveness of this technique in realistic situations.

In section 3.6 we consider the detection of sources with coded aperture imaging systems. The correlation image will contain peaks that are due to photon statistics as well as those due to the presence of sources. In the process of searching the image for sources, the largest of these noise peaks may be identified as sources. This problem was initially addressed by Dicke (1968) and later by Gunson and Polychronopoulos (1976) who presented approximations in the context of a discretely binned image. For the continuous correlation image we examine the distribution of peak heights in the absence of sources. We then present curves for determining the probability that a peak is due to noise, and apply these to a simple detection problem.

### 3.2 The Instrument Model

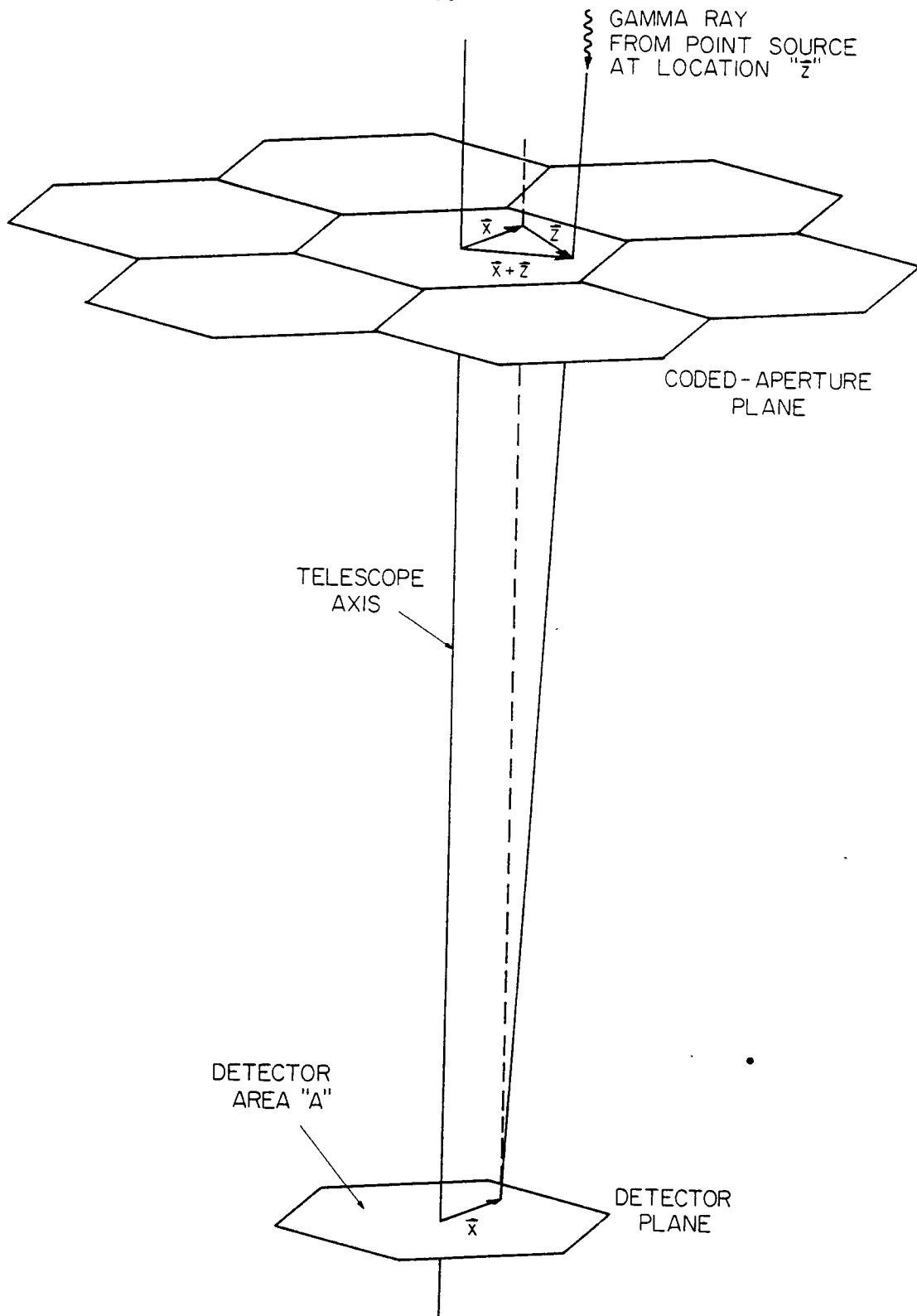
The basic elements of a coded aperture imaging instrument are a position-sensitive photon detector and a coded aperture mask. We will assume that the mask and detector lie in parallel planes and have a common central axis. We will also assume that the thickness of the mask and detector are much smaller than their separation. The detector will be assumed to have an area  $A$  that can precisely contain one repetition of the mask pattern. The assumed geometry is shown in Figure 3.1, with the hexagonal areas on the mask plan representing a single repetition of the mask pattern.

We will describe the mask by the function  $M(\vec{y})$  where  $\vec{y}$  is the position on the mask plane.  $M(\vec{y})$  is one if the mask is transparent at that position and zero if it is opaque. The flux incident on the detector due to a single point source is:

$$I(\vec{x}) = M(\vec{x} + \vec{z})F , \quad (3.1)$$

where  $\vec{z}$  is a planar vector describing the source location and  $F$  is the unmodulated source flux. As depicted in Figure 3.2,  $\vec{z}$  is tangential to the mask plane, with its length equal to the mask-detector separation times the tangent of the angle separating the source from the instrument's axis. We will restrict ourselves in this chapter to mono-energetic sources, although the extension to an energy spectrum is straightforward. The detector itself has a finite position resolution, which we will describe by the point spread function  $p(\vec{\Delta})$  which gives the probability of detecting a photon's incident position with error  $\vec{\Delta}$ . The actual detected flux distribution is the source flux distribution convolved with the point spread function, plus the detector background  $B(\vec{x})$ .

**Figure 3.2.** The geometry of a coded aperture instrument. The coded aperture or mask lies in a plane parallel to the detector with the mask and detector both centered on the telescope axis. The mask pattern is periodic, with a unit cell or single repetition of the pattern being represented in the figure by a hexagon. The orientation and size of the detector is the same as the unit cell of the mask pattern. A planar source coordinate  $\vec{z}$  is defined such that a photon from a source in direction  $\vec{z}$  that is incident on the detector plane at position  $\vec{z}$  must have passed through the mask plane at position  $\vec{z} + \vec{z}$ .



This may be expressed as

$$D(\vec{x}) = \sum_j M^{eff}(\vec{x} + \vec{z}_j) F_j + B(\vec{x}) , \quad (3.2)$$

where we have assumed that there are multiple point sources. The effective mask function is given by

$$M^{eff}(\vec{y}) = \int p(\vec{\Delta}) M(\vec{y} - \vec{\Delta}) d^2\Delta . \quad (3.3)$$

We now distinguish the mask pattern,  $M_+(\vec{y})$ , from the anti-mask pattern,  $M_-(\vec{y})$ . Similarly, we define  $D_+(\vec{x})$  and  $M_\pm^{eff}(\vec{y})$  as the mask and anti-mask representations of equations (3.2) and (3.3) respectively. The difference between the fluxes detected using the two masks contains the background independent information about the sources. This is given by

$$\Delta D(\vec{x}) = D_+(\vec{x}) - D_-(\vec{x}) = \sum_j \Delta M^{eff}(\vec{x} + \vec{z}_j) F_j \quad (3.4)$$

where

$$\Delta M^{eff}(\vec{y}) = M_+^{eff}(\vec{y}) - M_-^{eff}(\vec{y}) . \quad (3.5)$$

### 3.3 Source Flux and Direction Estimation

The optimal technique for estimating source fluxes and directions from the detected flux distributions is the maximum likelihood method. The likelihood function  $L$  gives the probability of the measured data being observed for a given model of the background distribution and source configuration. In the maximum likelihood method the source configuration and the background

distribution are estimated by the model that maximizes the likelihood function. Direct application of this method to the observations we are considering would be difficult because of the large number of events involved ( $10^6-10^8$ ). Fortunately however there is a simple approximation to  $L$  that is valid for the prolonged low source flux to background flux ratio observations in which we are interested. We will argue that for large number of events  $L$  tends toward a Gaussian, and derive this Gaussian form. From this form will emerge the correlation image, which summarized the information in the data needed to estimate the source parameters of any configuration of sources.

**The Likelihood Function.** We will consider an observation with equal mask and anti-mask measurements each of duration  $\tau/2$ . The observed data consists of a set of  $n_+$  positions  $\vec{x}_k^+$  of events detected during the mask observation and a set of  $n_-$  positions  $\vec{x}_k^-$  of events detected during the anti-mask observation. A model of the background distribution and source distribution will predict detector distributions  $\hat{D}_+(\vec{x})$  and  $\hat{D}_-(\vec{x})$  for the mask and anti-mask observations respectively. These may be expressed in terms of the model of the background and a set of source directions and locations through equation 4.2. The likelihood of the model is the probability of observing the events  $\vec{x}_k^+$  and  $\vec{x}_k^-$  assuming that the model is correct. Since the mask and anti-mask observations are conducted independently, the likelihood will be the product of the probabilities  $L_+$  and  $L_-$  of each observation. We will consider then the likelihood  $L_+$  of the mask observation. Given the model distribution  $\hat{D}_+(\vec{x})$  the mean number of events expected to be observed is given by

$$N_+ = \frac{\tau}{2} \int_A \hat{D}_+(\vec{x}) d^2x \quad (3.6)$$

where the integral is over the area  $A$  of the detector. The probability of observing  $n_+$  events is then given by the Poisson distribution

$$P_n(n_+) = \exp(-N_+) \frac{N_+^{n_+}}{n_+!} \quad (3.7)$$

The probability for a given event to have a position  $\bar{x}_k^+$  is given by

$$P_x(x_k^+) = \frac{\tau \hat{D}_+(\bar{x}_k^+)}{2AN_+} \quad (3.8)$$

Thus the probability of the complete mask observation is given by

$$L^+ = P_n(n_+) \prod_{k=1}^{n_+} P_x(\bar{x}_k^+) \quad (3.9)$$

or upon substitution

$$\ln L^+ = \sum_{k=1}^{n_+} \ln \hat{D}_+(\bar{x}_k^+) - \frac{\tau}{2} \int_A \hat{D}_+(\bar{x}) d^2x + \text{const.}$$

where the constant term depends only on the observation and not the model. When this is combined with a parallel result for the anti-mask observation we find that the likelihood function is given by

$$\ln L = \sum_{k=1}^{n_+} \ln \hat{D}_+(\bar{x}_k^+) + \sum_{k=1}^{n_-} \ln \hat{D}_-(\bar{x}_k^-) - \frac{\tau}{2} \int_A [\hat{D}_+(\bar{x}) + \hat{D}_-(\bar{x})] d^2x + \text{const.} \quad (3.10)$$

If we introduce the measured distribution functions

$$\rho_+(\bar{x}) = \frac{2}{\tau} \sum_{k=1}^{n_+} \delta(\bar{x} - \bar{x}_k^+) \quad \text{and} \quad \rho_-(\bar{x}) = \frac{2}{\tau} \sum_{k=1}^{n_-} \delta(\bar{x} - \bar{x}_k^-) \quad (3.11)$$



and let the index  $\alpha$  range over  $\pm 1$  then  $L$  may be expressed as

$$\ln L = \frac{\tau}{2} \sum_{\alpha} \int_A \left[ \rho_{\alpha}(\bar{x}) \ln \hat{D}_{\alpha}(\bar{x}) - \hat{D}_{\alpha}(\bar{x}) \right] d^2x + \text{const.} \quad (3.12)$$

**Gaussian Approximation for the Likelihood Function.** Due to the finite position resolution of the detector, any reasonable model  $\hat{D}_{\alpha}(\bar{x})$  of the detector flux distributions may be described by a fairly small number of parameters. In particular we may choose  $\hat{D}_{\alpha}(\bar{x})$  to be a linear combination of functions

$$\hat{D}_{\alpha}(\bar{x}) = \sum_{i=1}^N \hat{d}_i D_{\alpha}^i(\bar{x}) \quad (3.13)$$

The basis functions  $D_{\alpha}^i(\bar{x})$  may for example be chosen as combinations of spatial Fourier components. Then the finite position resolution of the detector limits the number of spatial frequencies required to accurately represent the detector distributions. This number will be proportional to

$$N \propto \frac{A}{\sigma^2} \quad (3.14)$$

where  $A$  is the detector area and  $\sigma$  is the width of the point spread function. When the number of events detected is much larger than the number  $N$  of coefficients  $\hat{d}_i$ , the likelihood function may be accurately described by a Gaussian in these coefficients.

We can derive the Gaussian form of  $L$  by evaluating the first and second derivatives of  $\ln L$  at the mean values of the estimated coefficients. These means  $d_i$  of  $\hat{d}_i$  correspond through (3.13) to the model flux distribution  $\hat{D}_{\alpha}(\bar{x})$

being equal to the expected distribution  $D_\alpha(\bar{x})$ . The Gaussian approximation is then given by

$$\ln L = \text{const.} + \sum_i (\hat{d}_i - d_i) \frac{\partial \ln L}{\partial \hat{d}_i} + \frac{1}{2} \sum_i \sum_j (\hat{d}_i - d_i)(\hat{d}_j - d_j) \left\langle \frac{\partial^2 \ln L}{\partial \hat{d}_i \partial \hat{d}_j} \right\rangle \quad (3.15)$$

which may be expressed as

$$\ln L = \text{const.} - \frac{\tau}{4} \sum_{\alpha} \int \frac{[\hat{D}_\alpha(\bar{x})]^2 - 2\rho_\alpha(\bar{x})\hat{D}_\alpha(\bar{x})}{D_\alpha(\bar{x})} d^2x \quad (3.16)$$

independent of the choice of functional basis.

The numerator in the integral in (3.16) can be seen to be the first two terms in the expansion of the square of the difference between the estimated and the measured flux distributions. Due to the finite position resolution of the detector, the model flux distribution and the expected flux distribution are limited in spatial frequency. The high-frequency components of  $\rho_\alpha(\bar{x})$  will therefore have no affect on the value of the likelihood function  $L$ . If these high-frequency components were filtered out, the square could be completed. This square difference is weighted with the expected mean distribution, whose exact form is unknown. However any reasonable approximation to this distribution will result in a reasonable approximation to  $L$  since  $L$  depends mainly on the closeness of the model distribution to the low-frequency portion of the measured distribution. We will approximate the expected mean distribution with a constant rate  $R$ , corresponding to the null case of a uniform background with no sources present.

**The Correlation Image.** At this point we can consider the estimation of the background distribution. We will introduce the differences  $\Delta\hat{D}(\bar{x})$  and

$\Delta\rho(\bar{x})$  of the estimated and the measured flux distributions between the mask and anti-mask observations. We will also introduce the corresponding sums  $\Sigma\hat{D}(\bar{x})$  and  $\Sigma\rho(\bar{x})$ . Then we have

$$\begin{aligned} \ln L = \text{const.} - \frac{\tau}{8R} \int \left( [\Sigma\hat{D}(\bar{x})]^2 - \Sigma\hat{D}(\bar{x})\Sigma\rho(\bar{x}) \right) d^2x \\ - \frac{\tau}{8R} \int \left( [\Delta\hat{D}(\bar{x})]^2 - \Delta\hat{D}(\bar{x})\Delta\rho(\bar{x}) \right) d^2x \end{aligned} \quad (3.17)$$

The background appears only in the sum of the estimated distributions and not in the difference. Thus estimating the background for a fixed source model involves only the first integral. However for any change in the source model we can find a background model that will produce the same sum distribution. Therefore the optimal value of the first integral is independent of the source model.

Estimation of source parameters then depends only on the second integral. If we introduce the model for the flux distribution given by equation (3.4) we have for the likelihood function after elimination of the background model

$$\ln L = \text{const.} - \frac{1}{2}\chi^2 \quad (3.18)$$

where

$$\chi^2 = \frac{A\tau}{4R} \sum_{jk} f(\bar{z}_j, \bar{z}_k) F_j F_k - \frac{1}{R} \sum_j C(\bar{z}_j) F_j \quad .$$

The correlation image function  $C(\bar{z})$  is a compact representation of the data that contain all the information relevant to estimating source parameters for

any source configuration. It is given by

$$C(\bar{z}) = \sum_k \alpha_k \Delta M^{eff}(\bar{z}_k + \bar{z}) \quad (3.19)$$

where  $k$  indexes events,  $\bar{z}_k$  is the location of an event, and  $\alpha_k$  is one or minus one depending on whether the event occurred with the coded aperture in the "mask" or "anti-mask" orientations.

The image function  $f(\bar{z}, \bar{z}')$  contains all of the information needed about the mask structure, and gives the expected contribution to the correlation image due to each point source. This is given by

$$f(\bar{z}, \bar{z}') = \frac{1}{A} \int_A \Delta M^{eff}(\bar{x} + \bar{z}) \Delta M^{eff}(\bar{x} + \bar{z}') d^2x \quad (3.20)$$

The image function takes on a very simple form when the masks are based on a Hexagonal Uniformly Redundant Array. As is described in the appendix, the image function depends only on the difference  $\bar{z} - \bar{z}'$  between the source locations, and will hereafter be noted as  $f(\bar{z} - \bar{z}')$ . The image function has a single peak at  $\bar{z} - \bar{z}' = \bar{\theta}$  surrounded by a flat region. In the case of well-separated sources this simple form of the image function permits an easy solution for the maximum likelihood estimates of the source fluxes and locations. In particular, the source locations are determined by setting the derivatives of  $\chi^2$  with respect to these locations equal to zero:

$$\bar{\theta} = \frac{\partial}{\partial \bar{z}} \left[ \frac{A\tau}{2} \sum_j f(\bar{z} - \bar{z}_j) F_j - C(\bar{z}) \right] \text{ at } \bar{z} = \bar{z}_k \text{ for } 1 \leq k \leq N \quad (3.21)$$

For well separated sources (  $f(\bar{z}_k - \bar{z}_j)$  small for  $k \neq j$  ) this reduces to the requirement that the minimum  $\chi^2$  estimates of the source locations be at

peaks in the correlation image. The source fluxes are determined by setting the derivatives of  $\chi^2$  with respect to these fluxes equal to zero:

$$0 = \left[ \frac{A\tau}{2} \sum_j f(\bar{z} - \bar{z}_j) F_j - C(\bar{z}) \right] \text{ at } \bar{z} = \bar{z}_k \text{ for } 1 \leq k \leq N . \quad (3.22)$$

If the sources are well separated the interdependence of the fluxes is small for all but the lowest order HURAs. This interdependence results from the fact that a uniform source distribution contributes a constant value to the detected flux and therefore can not be separated from the background.

### 3.4 Source Flux and Location Errors

We have seen that the correlation image arises from a maximum likelihood analysis. We could evaluate the source flux and location errors completely within this formalism, but an equivalent approach is to estimate the errors directly from the form of the correlation image  $C(\bar{z})$ . Proceeding from equations (3.4) and (3.19) we find that the expected value of the image is

$$\langle C(\bar{z}) \rangle = \frac{A\tau}{2} \sum_j f(\bar{z} - \bar{z}_j) F_j . \quad (3.23)$$

Note that the image function  $f(\bar{z})$  is simply the image expected from a point source. The image function also determines the covariance of the correlation image at two source locations:

$$\langle C(\bar{z}) C(\bar{z}') \rangle - \langle C(\bar{z}) \rangle \langle C(\bar{z}') \rangle \quad (3.24)$$

$$= \frac{\tau}{2} \int_A [D_+(\bar{x}) + D_-(\bar{x})] \Delta M^{eff}(\bar{x} + \bar{z}) \Delta M^{eff}(\bar{x} + \bar{z}') d^2x$$

$$\approx A\tau R f(\bar{z} - \bar{z}')$$

where  $R$  is the mean total count rate per unit detector area. The approximation assumes that the background is uniform, and that the mask plus anti-mask functions add to a constant.

Because the image function  $f(\bar{z})$  determines both the expected value and the covariance of the correlation image, it is important to understand its form. Starting from equation (3.20) and using the uniform redundancy and symmetry properties of the mask, the image function may be expressed simply as:

$$f(\bar{z}) = \frac{v}{A} \int_A h^{eff}(\bar{x}) h^{eff}(\bar{x} + \bar{z}) d^2x - \frac{1}{v} \quad (3.25)$$

where

$$h^{eff}(\bar{y}) = \int p(\bar{\Delta}) h(\bar{y} - \bar{\Delta}) d^2\Delta \quad . \quad (3.26)$$

The hexagonal cell function  $h(\bar{y})$  is defined to be periodic and have a value of one inside the central cell and its repetitions and zero elsewhere. Details of the derivation of equation (3.25) are given in the appendix. Note that the correlation image for an HURA is the same (but with better statistics) as that obtained with a mask with a single hexagonal hole.

We now estimate the error in determination of the source flux. The flux from a single point source is estimated from equation (3.22) to be:

$$F = \frac{2C(\bar{z}_{\max})}{A\tau f(\bar{\theta})} \quad (3.27)$$

where  $C(\bar{z}_{\max})$  is the maximum value of the correlation image. From equation (3.24), this flux estimate has an associated error

$$\sigma_F = \sqrt{\frac{4R}{A\tau f(\bar{\theta})}} \quad (3.28)$$

and the statistical significance of a source detection is given by

$$\kappa_F = \frac{F}{\sigma_F} = \kappa_0 \sqrt{f(\bar{\theta})} \quad \text{where} \quad \kappa_0 = \sqrt{\frac{A\tau F^2}{4R}}. \quad (3.29)$$

Here,  $\kappa_0$  is the statistical significance that would be obtained for a conventional on-source, off-source measurement with an instrument of the same effective area and the same observation time. For suitably designed coded aperture systems  $f(\bar{\theta})$  is close to unity, yielding close to optimum sensitivity.

In the case of multiple sources, the flux estimates and errors may be obtained from equations (3.22) and (3.24). If the number of sources is small compared to the number of cells in the HURA pattern, and if the sources are well separated, equations (3.27)-(3.29) are good approximations when  $C(\bar{z}_{\max})$  is understood as a relative maximum.

We now turn toward estimating the error in the determination of the source direction for the case of a single point source. If the displacement  $\Delta\bar{z}$  of the image peak from the true source location is small, then it can be determined from

$$0 = \left( \frac{\partial C}{\partial z_i} \right)_{\text{peak}} = \frac{\partial C}{\partial z_i} + \sum_j \frac{\partial^2 C}{\partial z_i \partial z_j} \Delta z_j \quad (3.30)$$

where the derivatives are to be evaluated at the true source location. The

mean value of the first derivatives are zero, while the second derivatives have

$$\langle \frac{\partial^2 C(\bar{z})}{\partial z_i \partial z_j} \rangle = \delta_{ij} \frac{A \tau F}{4} \nabla^2 f(\bar{\theta}) \quad (3.31)$$

as can be derived from equation (3.23) and the hexagonal symmetry of  $f(\bar{z})$ . We will assume that the statistical variations in the second derivatives are small compared to the magnitude of the diagonal elements. Thus, we will approximate the second derivatives by their expected value, and estimate that

$$\langle \Delta z_i \Delta z_j \rangle = \left[ \frac{A \tau F}{4} \nabla^2 f(\bar{\theta}) \right]^{-2} \langle \frac{\partial C}{\partial z_i} \frac{\partial C}{\partial z_j} \rangle = \frac{2}{\kappa_0^2 |\nabla^2 f(\bar{\theta})|} \delta_{ij} \quad (3.32)$$

The error in the source direction along any axis is then given by:

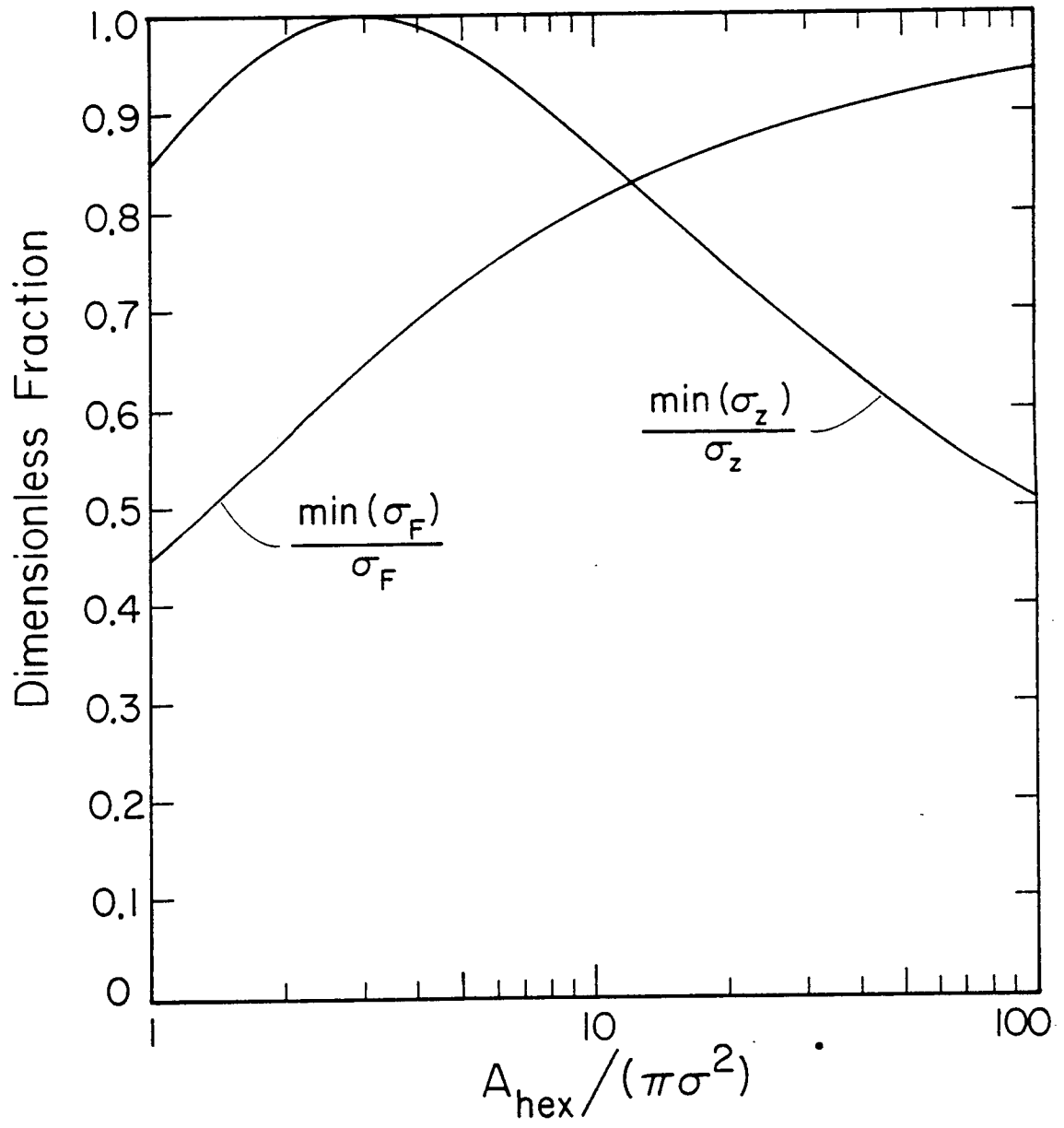
$$\sigma_s = \frac{1}{\kappa_0} \sqrt{\frac{2}{|\nabla^2 f(\bar{\theta})|}} \quad (3.33)$$

In the case of multiple sources, the source directions are uncorrelated if the sources are well separated, so that equation (3.33) is still valid.

We have obtained expressions for the errors on source fluxes and directions that are derived from the correlation image. These expressions depend only on the image function  $f(\bar{z})$ , which is fixed when the point spread function is known and when a width is chosen for the hexagonal cells. In Figure 3.3 we show the dependence of the flux error  $\sigma_F$  and the source location error  $\sigma_s$  on the choice of the hexagonal cell size. We assume that the point spread function is Gaussian with variance  $\sigma$ . Each curve compares the errors to the best obtainable from a measurement with the same source flux, observation



**Figure 3.3.** The dependence of the statistical flux measurement error  $\sigma_F$  and the source localization error  $\sigma_s$  on the choice of the mask cell size. The errors are compared to the minimum errors achievable with the same instrument by variation of the mask cell size. The size of the mask cell is parametrized in terms of the ratio of its area  $A_{\text{cell}}$  to  $\pi\sigma^2$  where  $\sigma$  is the width of the detectors (Gaussian) point spread function.



time, detector area and background rate. The ratio of errors is plotted versus the ratio of the hexagonal cell area,  $A_{\text{hex}} = A/v$ , to  $\pi\sigma^2$ . An approximation has been made that replaces the hexagon with a circle of the same area.

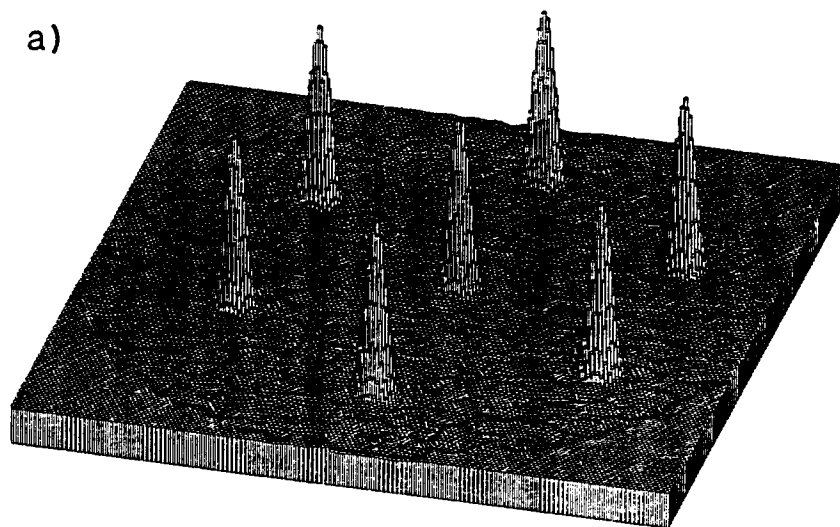
As can be seen from Figure 3.3, the source location error is optimized at small cell size, while the flux sensitivity is optimized in the limit of large cell size. The choice of cell size thus requires a trade-off between flux sensitivity and source localization ability. In practice, both good flux sensitivity and source localization capability can be achieved simultaneously. For example, with a cell size  $A_{\text{hex}}/\pi\sigma^2 = 10$ , the flux sensitivity is within 20% of the optimal sensitivity, while the source localization capability is within 15% of the optimum value.

### 3.5 Continuous Mask Rotation

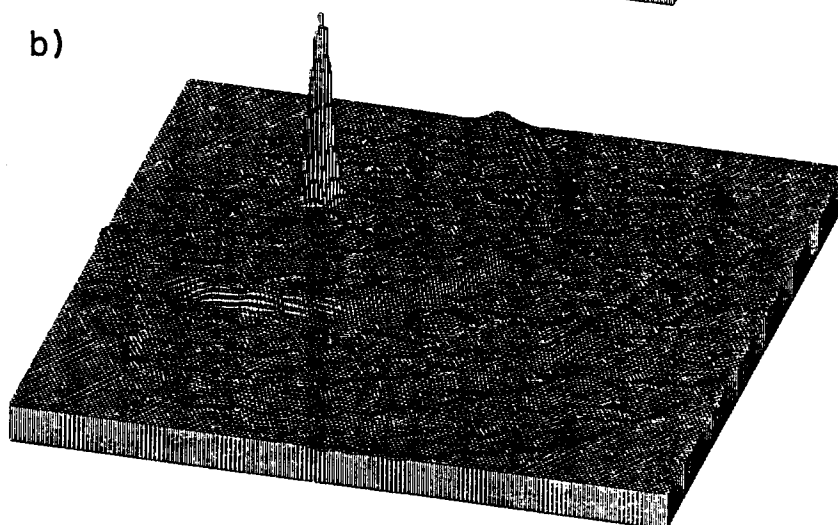
For simple mask-anti-mask imaging with a non-rotating periodic mask the image itself is periodic. Figure 3.4a shows one such image of a single source where the image has been calculated over a field-of-view containing several periods. The source location is ambiguous, the true source peak being repeated with the period of the mask. Continuous rotation of the mask removes this ambiguity allowing an extension of the field-of-view. With continuous mask rotation the data is first divided into sets corresponding to fixed mask orientation intervals, and these sets are then grouped into mask-anti-mask pairs. For each pair an image is formed as described above. The result is a series of images parametrized by mask rotation angle. The series of images contain the true source peak at a fixed location with the repeated peaks rotating about it. The result of adding the images together is shown in

**Figure 3.4.** Extension of the field-of-view by mask rotation. Figure 3.4a shows an image of a single source taken with a single mask orientation and its anti-mask orientation. The image is calculated over a wide field-of-view to reveal the inherent periodic ambiguity. Figure 3.4b shows an image of the same source taken with a rotating mask. The false peaks in 3.4a are reduced to rings by the mask rotation. These rings are removed by a cleaning procedure to produce the image in Figure 3.4c.

a)



b)



c)

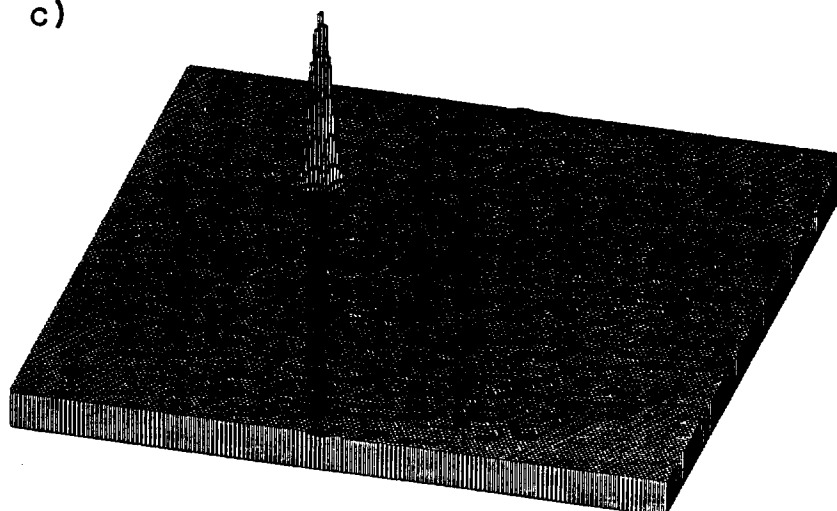


Figure 3.4b. The repeated peaks are blurred into a series of low rings centered on the peak at the true source location. Since the location and amplitude of the rings are directly related to the location and amplitude of the peak, the rings may be subtracted, resulting in the final image in Figure 3.4c.

This procedure of adding together the images from all mask rotation angles can be justified by an extension of the maximum likelihood analysis presented in section 3.3. For this analysis we must consider a continuous set of mask orientations parametrized by the mask orientation angle  $\phi$ . A model of the background and source configuration will predict a detector flux distribution for each mask orientation. However this angle-dependent flux distribution can still be described as a linear combination of a small number of functions. For example, we could employ as basis functions Fourier components that are stationary with the detector for the background, and Fourier components that rotate with the mask for the description of the modulated source flux. As a further extension, slow temporal variations in the background could be described with low-frequency temporal Fourier components. The number of coefficients needed for an accurate description of the detector flux distribution will still be small compared to the number of events in a typical astronomical observation. Thus the likelihood function can still be approximated with a Gaussian.

As long as temporal variations in the background are slow compared to the rotational period of the mask, the result of such an analysis is the definition of a new correlation image function  $C_{rot}(\bar{z})$  to replace  $C(\bar{z})$  and a new image function  $f_{rot}(\bar{z})$  to replace  $f(\bar{z})$ . The correlation image for a rotation mask observation is

$$C_{rot}(\bar{z}) = \sum_k \Delta M^{eff}(\bar{z}_k + \bar{z}, \phi_k) \quad (3.34)$$

where  $\Delta M^{eff}(\bar{z}, \phi)$  is  $\Delta M^{eff}(\bar{z})$  rotated with the mask, and  $\phi_k$  is the orientation angle of the mask at the time of event  $k$ . This is the image just described. The image function for a rotating mask observation is

$$f_{rot}(\bar{z}) = \frac{1}{2\pi} \int_0^{2\pi} f(\bar{z}, \phi) d\phi \quad (3.35)$$

With the substitution of  $C_{rot}(\bar{z})$  for  $C(\bar{z})$  and  $f_{rot}(\bar{z})$  for  $f(\bar{z})$  the principal equations and result of the analysis in section 3.3 and 3.4 still hold.

### 3.6 Source Detection and Noise Peaks

In section 3.3 and 3.4 we discuss the estimation of source fluxes and directions from the location and height of peaks in the correlation image. Implicit in this discussion was the assumption that the peaks being considered were due to the presence of photon sources. The correlation image will however also contain peaks that are due to photon statistical fluctuations. In this section we will examine the criteria for deciding that a peak is due to a source rather than a statistical fluctuation.

Throughout this discussion we will assume that the angular size of the mask cells have been chosen so that the fluctuations in the cosmic background discussed in chapter 2 are much smaller than the photon statistical uncertainty of flux measurements made by the instrument. For a typical field-of-view the expected correlation image will then consist largely of a nearly uniform constant level, with perhaps a few peaks due to any large sources

present. The image constructed from an actual observation will consist of this expected image with added photon statistical fluctuations which for any fixed point in the image will have a Gaussian distribution with variance given by equation (3.24). The process of searching the image for peaks will increase our chances of finding a large fluctuation, so that peaks due to statistical fluctuations may be found that are large compared to the root mean square fluctuation level.

We must stress that this bias toward large fluctuations produced by searching is not restricted to imaging instruments, but rather is a result of the resolution requirements imposed on the instrument. A survey of the same region of the sky with a non-imaging instrument with the same angular resolution would require the independent measurement of numerous fluxes. The same kind of bias would result from searching through these numerous measurements for large excesses.

**The Noise Peak Distribution.** To confidently decide that a peak in an image is due to a source we must know the probability of it occurring in the image due to a statistical fluctuation. We will therefore examine the expected distribution of peak heights in images made with no sources present. With no sources present a peak is equally likely to occur at any point in an image. By a peak we mean a local maxima, so that the presence of a peak near the point  $\bar{z}$  in a given image  $C(\bar{z})$  depends only on the first and second derivatives of the image at  $\bar{z}$ . Thus the probability of finding a peak near the point  $\bar{z}$  of a given height in any pure noise image should depend only on the joint probability of having the given magnitude and the proper first and second derivatives at this point. The image value and the image derivatives



are linear combinations of a large number of random variables. Thus, by the central limit theorem, the joint probability distribution will be Gaussian, and completely determined by the means and covariances of the image value and derivatives. With this distribution established we may then calculate the probability of finding a peak of a given height.

To find the distribution of noise peak heights then, our initial task is to establish the joint probability of obtaining a given image value and first and second derivatives at the point  $\bar{z}$ . We will first normalize the image to its expected variance :

$$\hat{C}(\bar{z}) = \frac{C(\bar{z})}{\sqrt{A \tau R f(\bar{\theta})}} \quad (3.36)$$

With this normalization our results should be independent of the observation time. For notational convenience we now define a 6-dimensional vector  $\bar{a}$  with components

$$a_0 = \hat{C}(\bar{z}) \quad (3.37)$$

$$a_1 = \frac{\partial \hat{C}(\bar{z})}{\partial z_x} \quad a_2 = \frac{\partial \hat{C}(\bar{z})}{\partial z_y}$$

$$a_3 = \frac{1}{2} \left[ \frac{\partial^2 \hat{C}(\bar{z})}{\partial z_x^2} + \frac{\partial^2 \hat{C}(\bar{z})}{\partial z_y^2} \right] \quad a_4 = \frac{1}{2} \left[ \frac{\partial^2 \hat{C}(\bar{z})}{\partial z_x^2} - \frac{\partial^2 \hat{C}(\bar{z})}{\partial z_y^2} \right] \quad a_5 = \frac{\partial^2 \hat{C}(\bar{z})}{\partial z_x \partial z_y} .$$

We know by the central limit theorem that in the limit of a large number of events the joint probability  $Q(\bar{a})$  tends toward the multivariate normal form

$$Q(\vec{a}) = \frac{\sqrt{\text{Det}(\eta)}}{(2\pi)^3} \exp(-\frac{1}{2} \sum_{ij} \eta_{ij} (a_i - \langle a_i \rangle)(a_j - \langle a_j \rangle)) \quad (3.38)$$

where the  $\eta_{ij}$  are the elements of  $\eta$ , the inverse of the covariance matrix,

$$V_{ij} = \langle a_i a_j \rangle - \langle a_i \rangle \langle a_j \rangle . \quad (3.39)$$

Since there are no sources present, equation (3.23) implies that the expectation value of  $\vec{a}$  is zero. The covariance matrix may be found with the help of equation (3.24), which for  $\langle \hat{C}(\vec{z}) \rangle = 0$  reduces to

$$\langle \hat{C}(\vec{z}) \hat{C}(\vec{z}') \rangle = \frac{f(\vec{z} - \vec{z}')}{f(\vec{0})} . \quad (3.40)$$

Thus for example

$$V_{11} = \frac{\partial^2}{\partial z_x \partial z_x} \langle \hat{C}(\vec{z}) \hat{C}(\vec{z}') \rangle |_{\vec{z}=\vec{z}'} = \frac{-1}{f(\vec{0})} \frac{\partial^2 f(\vec{0})}{\partial z_x^2} . \quad (3.41)$$

Due to the symmetry of the image function  $f(\vec{z})$  and the linear combinations chosen for the second derivatives, most of the off-diagonal elements of  $V_{ij}$  are zero. The only non-zero off-diagonal elements are due to an anti-correlation between the image value and the image curvature. The covariance matrix is

$$[V] = \begin{bmatrix} 1 & 0 & 0 & -\beta & 0 & 0 \\ 0 & \beta & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 & 0 & 0 \\ -\beta & 0 & 0 & 2\gamma & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma \end{bmatrix} \quad (3.42)$$

where

$$\beta = \frac{-1}{2f(\vec{0})} \nabla^2 f(\vec{0}) \quad \text{and} \quad \gamma = \frac{1}{8f(\vec{0})} \nabla^4 f(\vec{0}) . \quad (3.43)$$

The reduction of non-zero elements to two parameters is due to the symmetry of the image function, a reflection of the URA property of the mask and the hexagonal shape of its cells.

Having determined the covariances and through them the distribution  $Q(\bar{a})$ , our next task is to establish the conditions for the existence of a peak near  $\bar{z}$  given a set  $\bar{a}$  of image value and derivatives. By a peak we will mean a local maximum. At a peak the first derivatives are zero and the second derivatives are such that along any axis passing through the peak the second derivative is negative. The conditions on the second derivatives can be stated as follows :

$$0 > \text{Tr}\left(\frac{\partial^2 \hat{C}(\bar{z})}{\partial z_j \partial z_j}\right) = 2a_3 \quad 0 < \text{Det}\left(\frac{\partial^2 \hat{C}(\bar{z})}{\partial z_j \partial z_j}\right) = a_3^2 - a_4^2 - a_5^2 \quad . \quad (3.44)$$

If the conditions on the second derivatives are satisfied, the displacement  $\Delta \bar{z}$  of the nearest peak from the point  $\bar{z}$  may for small displacements be found from

$$0 = \left( \frac{\partial \hat{C}}{\partial z_i} \right)_{\text{peak}} = \frac{\partial \hat{C}}{\partial z_i} + \sum_j \frac{\partial^2 \hat{C}}{\partial z_j \partial z_j} \Delta z_j \quad (3.30)$$

where the derivatives are to be evaluated at  $\bar{z}$ . To find the probability  $P(a_0)dA$  that there is a peak of height  $a_0$  within an area  $dA$  of the point  $\bar{z}$ , we need to integrate the distribution  $Q(\bar{a})$  over those values of the first and second derivatives for which (3.44) holds and for which  $\Delta \bar{z}$  is within  $dA$ . In the limit of  $dA \rightarrow 0$ , the probability per unit area of a peak of given height is

$$P(a_0) = \int_{\substack{a_1^2 > a_4^2 + a_5^2 \\ a_3 < 0}} Q(\bar{a}) \delta(\Delta \bar{z}) da_1 da_2 da_3 da_4 da_5 \quad (3.46)$$

where  $\delta()$  denotes the Dirac delta function. From equation (3.45) we can deduce that

$$\delta(\Delta \bar{z}) = |Det(\frac{\partial^2 \hat{C}}{\partial z_j \partial z_j})| \delta(\nabla \hat{C}) = |a_3^2 - a_4^2 - a_5^2| \delta(a_1) \delta(a_2) \quad (3.47)$$

Upon substitution of equation (3.38) for  $Q(\bar{a})$  the integration is straightforward. Noting that the significance  $\kappa$  of a peak in the image is given by  $a_0$ , we find that the probability per unit area of a noise peak of significance  $\kappa$  is

$$P(\kappa) = \frac{F}{A_0} \exp(-\frac{1}{2} B \kappa^2) [G \exp(\frac{1}{2} G^2 \kappa^2) Z(G \kappa) + r^2 (\kappa + [\kappa^2 - 1] H \exp(\frac{1}{2} H^2 \kappa^2) Z(H \kappa))] \quad (3.48)$$

where

$$Z(x) = \int_{-\infty}^x \exp(-\frac{1}{2} y^2) dy$$

$$r = \frac{\beta}{\sqrt{2\gamma}}$$

$$F = \frac{\sqrt{3}}{\pi} \frac{\sqrt{1-r^2}}{r}$$

$$A_0 = 2\pi \sqrt{3} \frac{\beta}{\gamma}$$

$$B = \frac{1}{1-r^2}$$

$$G = \frac{r}{\sqrt{(3-2r^2)(1-r^2)}}$$

$$H = \frac{r}{\sqrt{1-r^2}} .$$

The area  $A_0$  contains on average one peak.

Figure 3.5 shows examples of these noise peak distributions. The various parameters have been calculated for a Gaussian point spread function with width  $\sigma$ , which is kept fixed while the width of the basic hexagonal mask cell is varied. In the calculation we approximated the hexagon with a circle of the same area. Each curve gives the number of peaks, per unit of peak significance, expected within an area  $\pi\sigma^2$  of the image. Each curve is labeled with the ratio of the hexagon cell area to  $\pi\sigma^2$ .

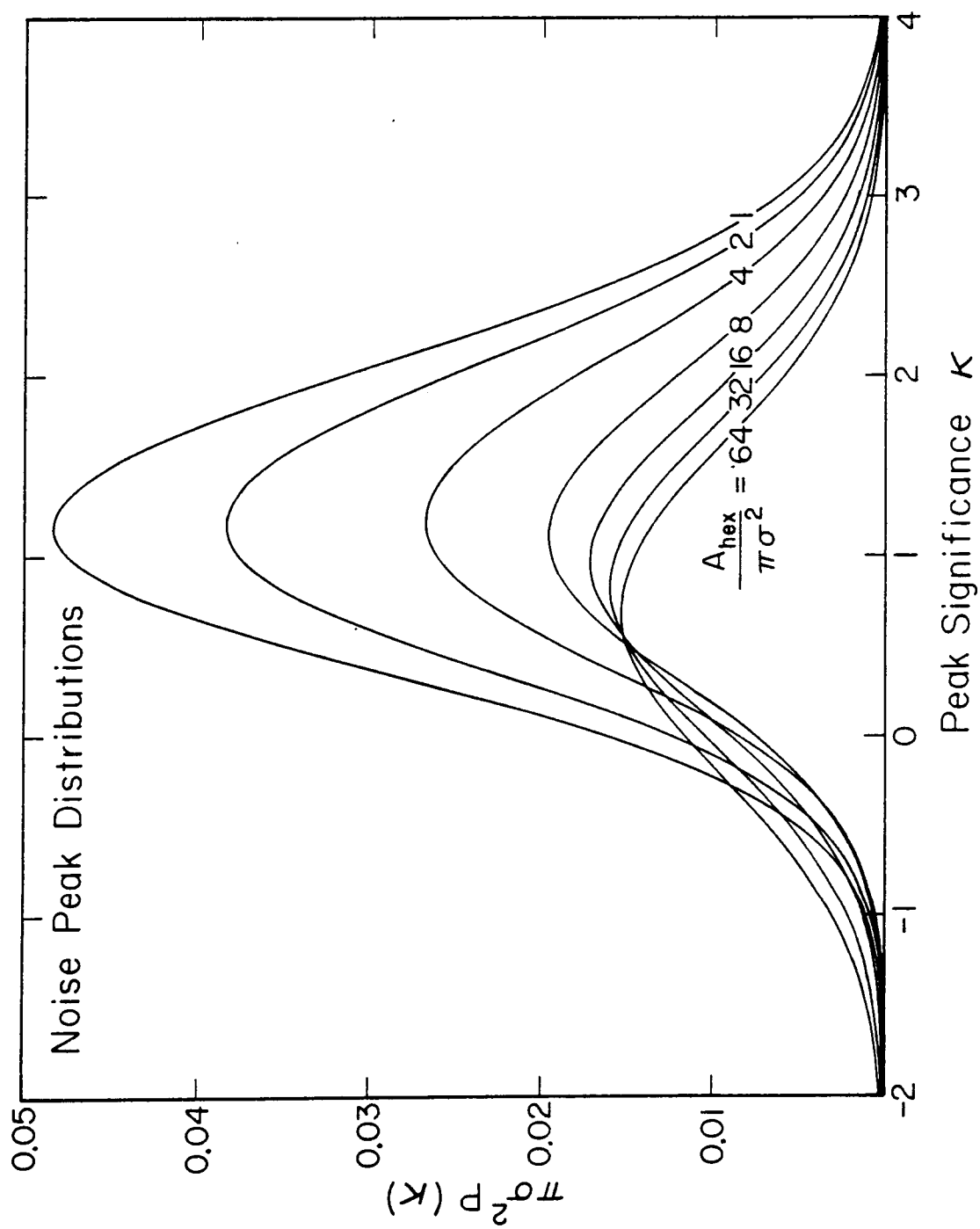
As can be seen from Figure 3.5, the masks made from smaller cells result in images with more noise peaks, and a distribution of peaks at higher significance. In the extreme limit of small hexagons, the total number of peaks becomes independent of the cell size, and depends only on the width of the point spread function.

In Figure 3.6 we show the integral of the distributions, again parameterized by the ratio of the mask cell to  $\pi\sigma^2$ . The integral distribution is defined by

$$P(>\kappa) = \int_{\kappa}^{\infty} P(\kappa) d\kappa \quad (3.49)$$

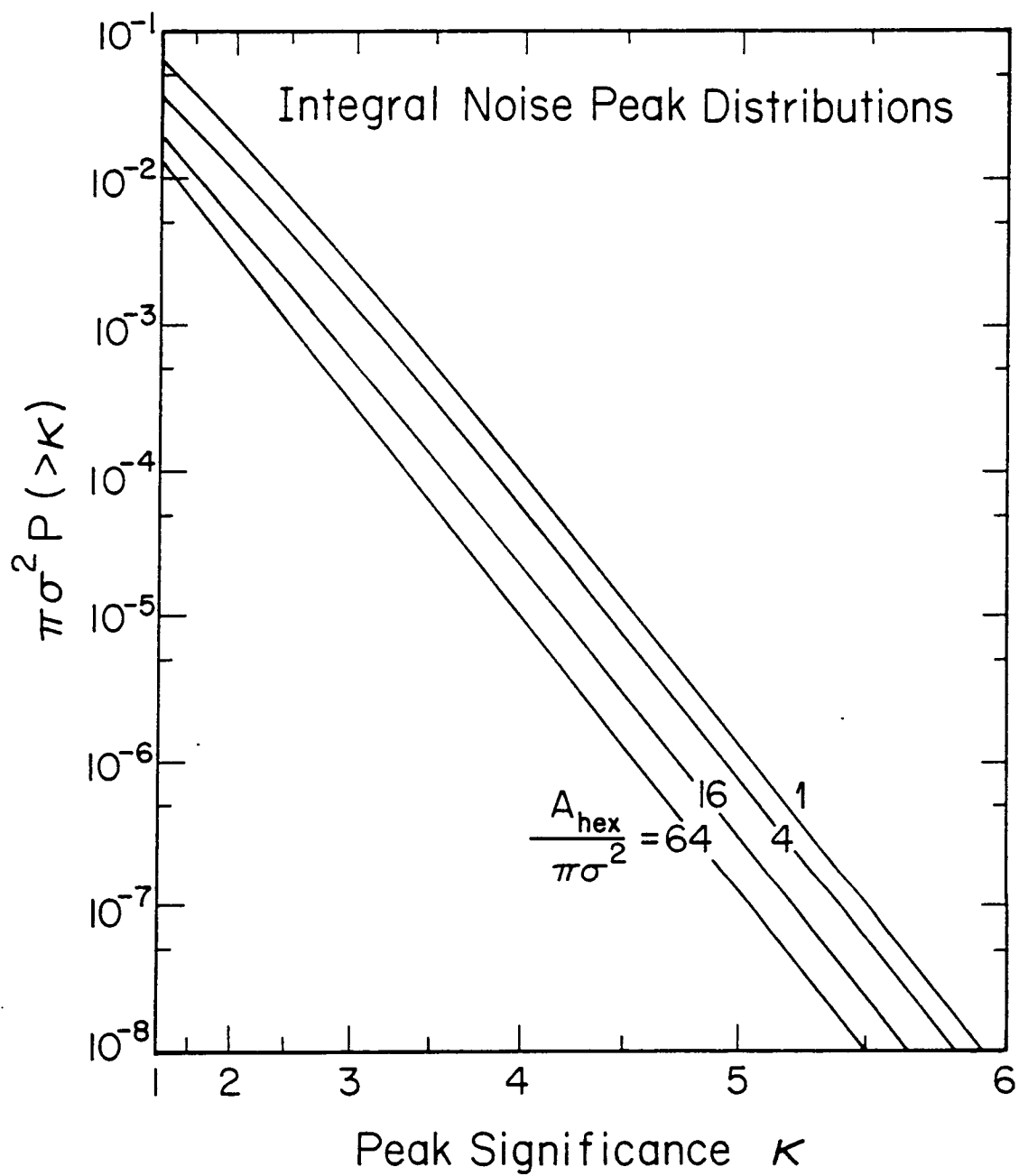
and give the total number of noise peaks expected per unit area above the significance  $\kappa$ . Each curve in Figure 3.6 gives the number of peaks with significance greater than  $\kappa$  expected within an area  $\pi\sigma^2$  of an image.

**Figure 3.5.** Noise peak distributions for various choices of cell size. Each curve is labeled by the ratio of cell area  $A_{\text{hex}}$  to detector resolution area  $\pi\sigma^2$  where sigma is the detector's (Gaussian) point spread function width. The curves give the number of peaks per unit significance expected within an area  $\pi\sigma^2$  with significance  $\kappa$ .



**Figure 3.6** Integral noise peak distributions for various choices of cell size. Each curve is labeled by the ratio of cell area  $A_{\text{cell}}$  to detector resolution area  $\pi\sigma^2$  where sigma is the detector's (Gaussian) point spread function width. The curves give the number of peaks expected in the area  $\pi\sigma^2$  above a significance  $\kappa$ . The  $\kappa$  scale is linear in  $\kappa^2$ .





The log of this number is roughly linear in  $\kappa^2$ , so we have chosen to plot  $\kappa$  along an axis linear in  $\kappa^2$ . As can be seen the distributions drop off quickly, with, for example, an image with an area of  $1000A_{hez}$  and  $A_{hez}/\pi\sigma^2 = 4$ , having on average more than one peak per image with significance greater than 3.5 but less than one peak per forty images with significance greater than 4.5.

**Single Source Detection.** To illustrate the use of these distributions we will consider a simple detection problem. We will suppose that one of two hypotheses is true; either the null hypothesis with no sources in the field-of-view, or the source hypothesis with a single source in the field-of-view with statistical significance  $\kappa_0$ . The problem then is, how can we decide between these two possibilities for a given observation, and how accurate is this decision making process? To address the accuracy of the decision making process we need to know two things, the probability of a correct decision given that a source is present and the probability of an incorrect decision given that no source is present. These probabilities are known respectively as the True Positive Fraction (TPF) and the False Positive Fraction (FPF).

A standard method for deciding between two hypotheses is to use the likelihood ratio test. The likelihood ratio is the ratio of the maximum value of the likelihood function under the source hypothesis to the maximum value of the likelihood function under the null hypothesis. The source hypothesis is chosen if the likelihood ratio is greater than a threshold value. The threshold used is chosen to optimize some combination of the TPF and the FPF. For this problem the likelihood ratio is a monotonic function of the maximum value of the correlation image. Thus we will choose between the source and null hypothesis by comparing the significance of the largest peak in the

correlation image to some threshold value  $\kappa_{th}$ .

We will be interested only in thresholds for which the false positive fraction is small. Then since the probability of a noise peak having significance greater than the threshold is small, it is reasonable to assume that no more than one noise peak in an image will be above this threshold. With this assumption we have

$$FPF = A_I P(>\kappa_{th}) \quad (3.50)$$

where  $A_I$  is the total area of the image being considered.

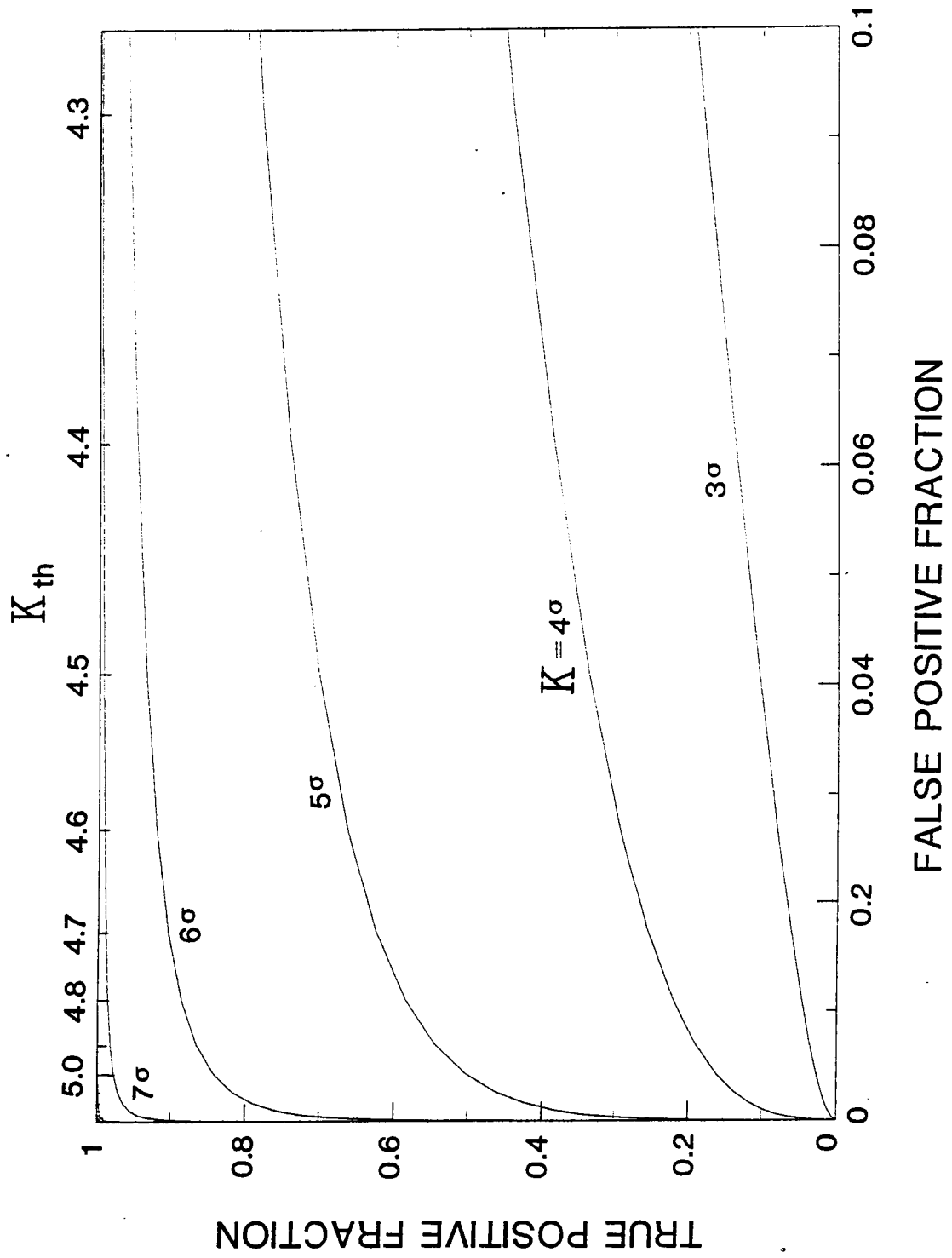
If a source is present and yet is not detected the significance of the image in the source direction must be below the threshold, and no noise peaks may be above the threshold. The probability of this happening is the False Negative Fraction and is given by

$$FNF = \frac{Z(\kappa_0 - \kappa_{th})}{\sqrt{2\pi}} [1 - A_I P(>\kappa_{th})] \quad (3.51)$$

from which we may calculate the true positive fraction from  $TPF = 1 - FNF$ .

Figure 3.7 shows the relationship between the TPF and the FPF for an image with  $A_{hez}/\pi\sigma^2 = 8$  and for a moderate size image with an area of  $1000A_{hez}$ . Each curve is labeled with the source significance  $\kappa_0$  with points along the curve corresponding to different values of the threshold  $\kappa_{th}$ . The top axis is labeled with the value of the threshold  $\kappa_{th}$ . If for example we require that only one out of twenty fields-of-view may in error be declared to contain a source, then a threshold of  $\kappa_{th} = 4.45$  is needed. With this threshold we can

**Figure 3.7.** The effect of image searching on source detection. An image is constructed from one of two possible source configurations: either no source is present, or a single source of statistical significance  $\kappa_0$  is present at some random location. A decision is made between these possibilities by comparing the significance of the largest peak in the image with a threshold  $\kappa_{th}$ . The figure shows the average outcome of the decision. Each curve is labeled with the value of  $\kappa_0$ . The curves show the relationship of the false positive fraction, the fraction of images with no sources that are identified as containing a source, to the true positive fraction, the fraction of images containing a source that are correctly identified. Different points on the same curve correspond to different threshold values  $\kappa_{th}$ .



correctly identify more than half of the source fields-of-view only for sources of significance greater than 4.5.

## 4. The Gamma-Ray Imaging Payload (GRIP)

### 4.1 Overview

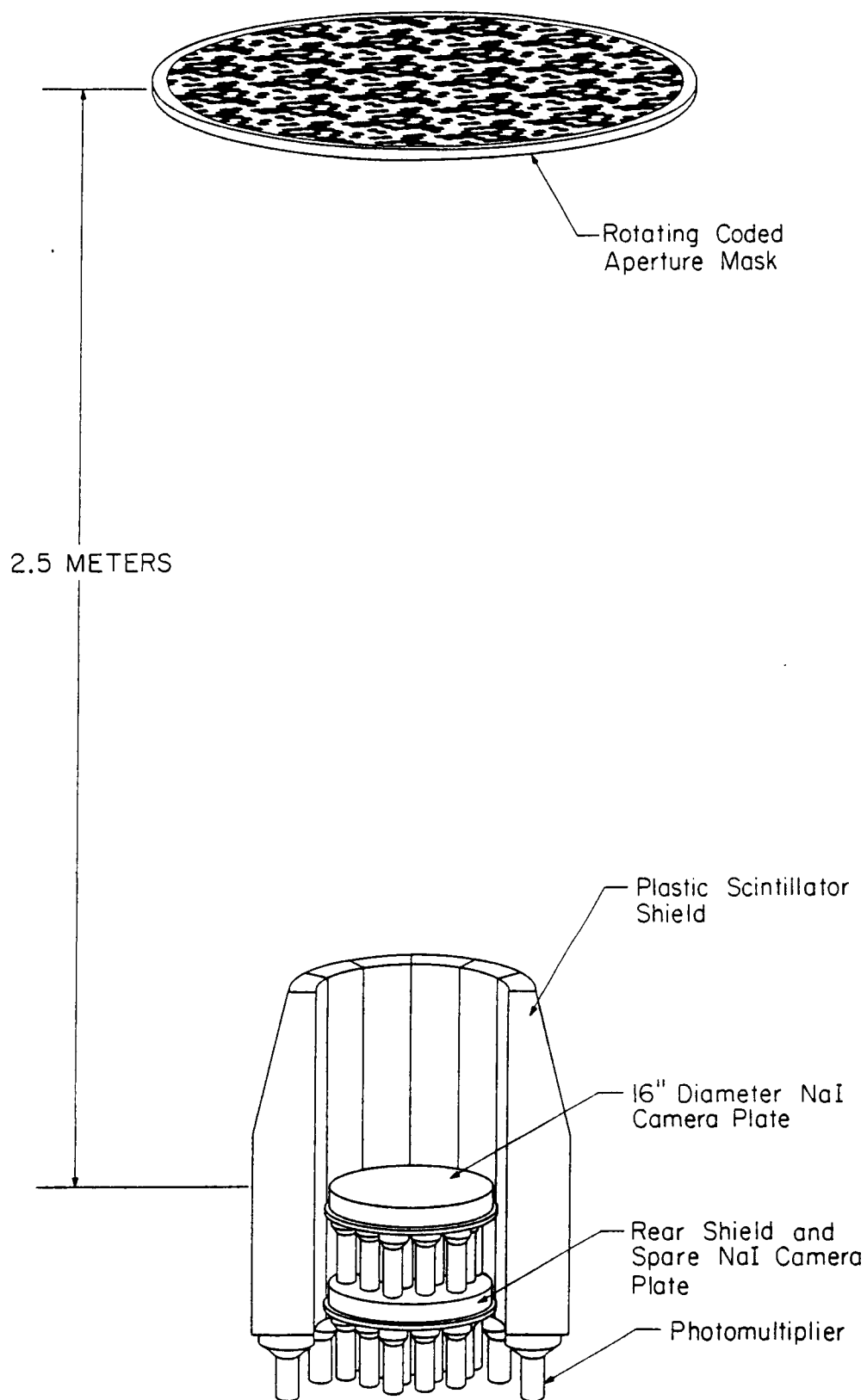
This chapter describes a balloon-borne coded aperture telescope for low-energy gamma-ray astronomy observations. This instrument, called GRIP (Gamma-Ray Imaging Payload) is designed for measurements in the energy range from 30 keV to 5 MeV with an angular resolution of  $0.6^\circ$  over a  $20^\circ$  diameter field-of-view. It employs a rotating hexagonal uniformly redundant array mask, implementing the imaging concepts developed in chapter 3.

The GRIP instrument has been designed for making significant observations as a high-altitude balloon payload. In addition it provides a proving ground for coded aperture imaging concepts. Although the GRIP instrument meets the angular resolution requirements set out in chapter 2, it is not capable of meeting the sensitivity objectives set out in chapter 1. This however is a matter of scale: a comparable instrument in a satellite environment could, with the longer exposures possible, meet these objectives.

The basic imaging elements of the GRIP instrument are shown in Figure 4.1. They consist of a shielded detector system separated by 2.5 meters from a lead coded aperture mask. The primary detector is a position-sensitive scintillator which detects the modulated flux of photons that pass through the mask. This is surrounded by active anti-coincidence shields to reduce the background from environmental photons.

**Figure 4.1.** Basic elements of the GRIP instrument. Shown are the rotating HURA mask and the shielded detector system. The primary NaI(Tl) detector is separated from the mask by 2.5 m. The primary detector is actively shielded by 12 plastic scintillator modules and a second NaI(Tl) detector.





Primary Detector	41 cm diameter $\times$ 5 cm NaI(Tl) Anger Camera Position Resolution: $< 5\text{mm rms}$ 0.1-5.0 MeV
Shield	Back Plate: 5 cm NaI(Tl) Sides: 16 cm plastic scintillator
Energy Range	0.03-5.0 MeV
Energy Resolution	8.3 keV FWHM @ 50 keV 70 keV FWHM @ 1 MeV
Mask	Hexagonal Uniformly Redundant Array Rotation rate: 1 rpm Cell Size: 2.54 cm flat-to-flat Mask-Detector Spacing: 2.5 m Size: 1.2 m diameter $\times$ 1.91 cm (Pb)
Imaging	Resolution: $0.6^\circ$ 1070 resolution elements in $20^\circ$ FOV Angular Localization: 3 arc min ( $10\sigma$ source)

**Table 4.1.** Characteristics of the GRIP balloon-borne gamma-ray telescope.

Table 4.1 summarizes the basic characteristics of the GRIP telescope. The primary detector is a NaI(Tl) camera plate manufactured by the Harshaw Chemical Company. This crystal is instrumented as an Anger camera, being viewed by 19 3-inch photo-multiplier tubes (PMTs) arranged in a hexagonal close-packed structure. The tubes are individually pulse height analyzed and the signals recorded for post-flight analysis.

The detector is surrounded by an anti-coincidence shield consisting of 12 plastic scintillator modules arranged in a cylinder and a NaI(Tl) crystal closing off the bottom of that cylinder. Each plastic shield piece is viewed by a 5 inch PMT. The lower NaI(Tl) crystal is identical to the primary detector, but only the analog sum of the PMT signals has been instrumented.

The mask is made of cast lead hexagons supported by an aluminum honeycomb sandwich which is transparent at gamma-ray energies. These lead

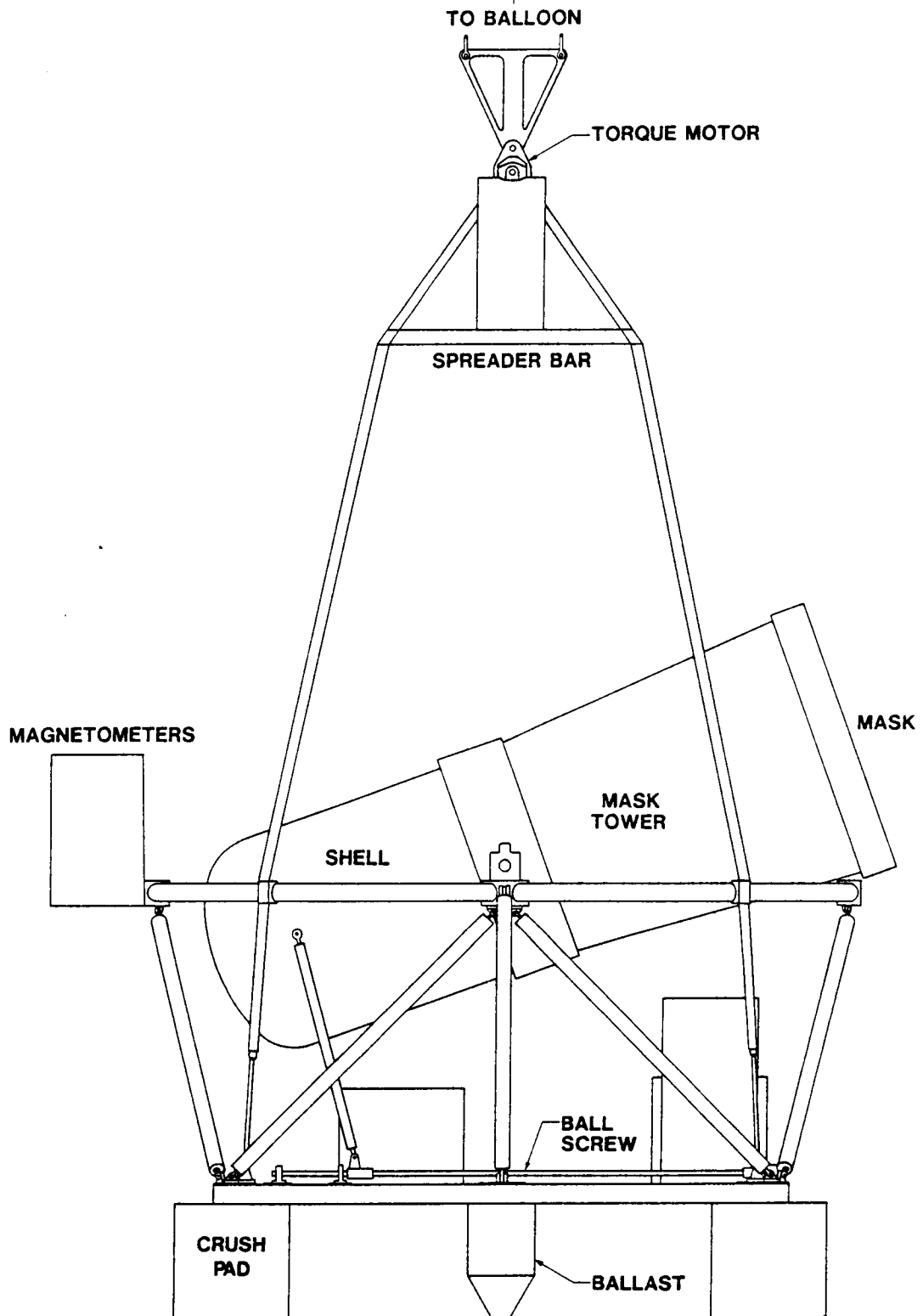
cells are arranged in the half-open, half-closed pattern of a hexagonal uniformly redundant array. Continuous rotation of the mask provides a temporal modulation of the flux. Due to the anti-symmetry of the mask on rotation by  $60^\circ$ , a position-by-position background subtraction can be performed every  $120^\circ$  of rotation. This rotation also removes the periodic ambiguity inherent in uniformly redundant arrays.

The shielded detector system and associated electronics are mounted in a pressure vessel often referred to as "the shell." As is shown in Figure 4.2, the mask is supported by a conical tower which is rigidly attached to the shell. The shell is attached to a platform by two bearings which allows changes in the elevation angle of the telescope. The elevation angle is controlled by a ball screw drive located on the platform. Also mounted on the platform are batteries and packages of electronics for pointing control, telemetry, and data recording. The platform is suspended by Kevlar straps from a spreader bar attached to a torque motor. In flight the torque motor produces torques between the helium balloon and the platform to control the instrument's azimuthal orientation. Azimuth stabilization and orientation are achieved using active magnetometer feedback to the azimuthal torque motor.

For each event in the primary detector with valid coincidence, all nineteen 12-bit primary detector PMT pulse heights, along with coincidence and timing information, are recorded. Event rates of up to  $5 \times 10^3$  events/s are possible, requiring a recording system with a 1 Mbit/s data rate. A 1.4 Mbit/s recording system with a total capacity of 25 Gbytes was developed for GRIP, using commercial VCR's and audio digitizers.

**Figure 4.2.** The GRIP instrument showing its configuration during flight. The primary detector and its shields are contained within a pressure vessel called the shell. The mask is supported by a conical tower which is attached to the shell. The orientable telescope, consisting of the shell, the mask tower and the mask, are attached to a pointing platform by two elevation bearings. The elevation of the telescope is controlled by a ball-screw drive. The platform is suspended by straps from a spreader bar attached to a torque motor. The azimuth of the telescope is controlled actively by the torque motor, using feedback from magnetometers for stabilization. Ballast is used for altitude control, while crush pads minimize damage upon landing.

- 133 -



The GRIP instrument was ready for its initial flight from Palestine, Texas in the fall of 1985. However problems with surface weather conditions and balloon materials prevented a flight at that time. GRIP was successfully flown in the fall of 1986, performing 24 hours of high-altitude gamma-ray observations. The data from these observations are presently in the initial stages of analysis.

The discussion of GRIP that follows will emphasize the components of primary scientific interest. In section 4.2 we will consider the main components of the imaging system, the mask and the position-sensitive detector with its active shield. In Section 4.3 we will describe the sub-systems that support the imaging system.

## **4.2 The Imaging System**

The essential components of a coded aperture gamma-ray imaging telescope are a coded aperture mask and a position-sensitive gamma-ray detector. The analysis of coded aperture imaging in chapter III employed a simple model of these elements. In this section we consider the details of the design and instrumentation of these components for the GRIP telescope. We will begin by discussing the design and construction of the mask. We then turn to the design and testing of the position-sensitive detector and its position algorithm. Then after a brief description of the on-board system for calibrating the detector, we conclude by discussing the active shielding of the primary detector.

#### 4.2.1 The Mask

Figure 4.3 shows the pattern of the mask constructed for the GRIP instrument. The mask is based on a Hexagon Uniformly Redundant Array pattern of order 79. Each closed cell consists of a cast hexagonal lead block which is 2.54 cm flat-to-flat and 1.91 cm thick.

As was discussed in chapter 3, the choice of the mask cell size is related to the position resolution of the detector, and involves a trade-off between the telescope's flux sensitivity and its source localization accuracy. With the cell size chosen, the source localization accuracy is within 26% of optimal, and the flux sensitivity within 32% of optimal, for the 50 keV to 2 MeV energy range.

The choice of the lead thickness was a compromise between gamma-ray opacity and the total weight of the mask. For a balloon payload excess mass results in a reduction of the obtainable altitude, and therefore a reduction in the source flux due to atmospheric attenuation. As is shown in Figure 4.4, with the thickness chosen the opacity is better than 90% below 700 keV and reaches a minimum of 60% near 3 MeV.

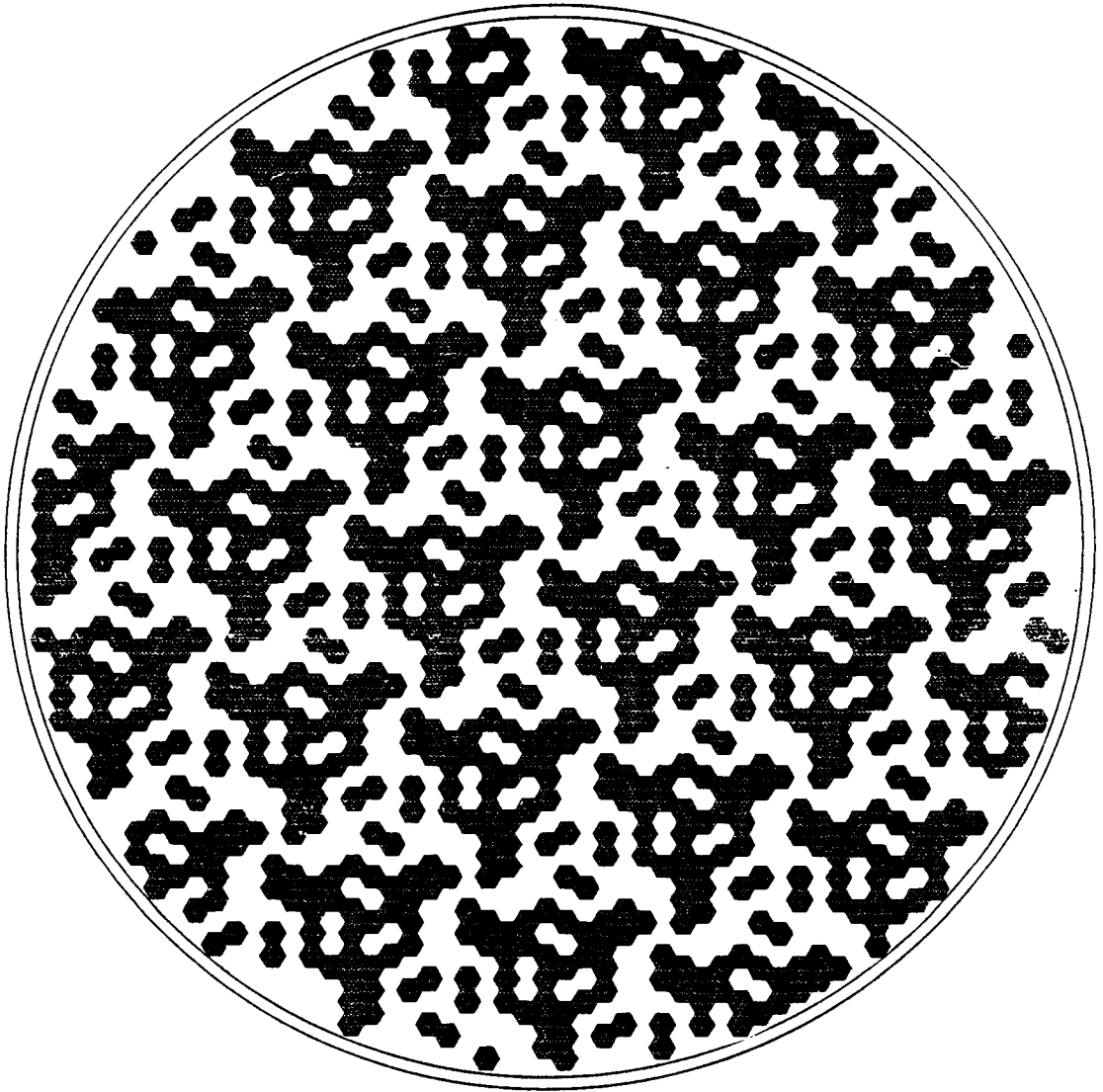
To the bottom of each lead block is bonded a 0.8 mm layer of tin which provides 3.8 attenuation lengths to suppress the 80 keV fluorescence x-rays produced in the lead. The blocks are mounted on a one-inch thick aluminum honeycomb sandwich bonded at the edges to a 117 cm diameter aluminum ring. The full assembly contains over 1000 lead blocks and weighs 142 kg.

The separation between the mask and the primary detector is 2.50 m. The angular resolution of the instrument is determined by the ratio of the mask cell size to this separation, with the chosen separation giving a resolution of 0.6 degrees. We considered larger separations to be impractical for a

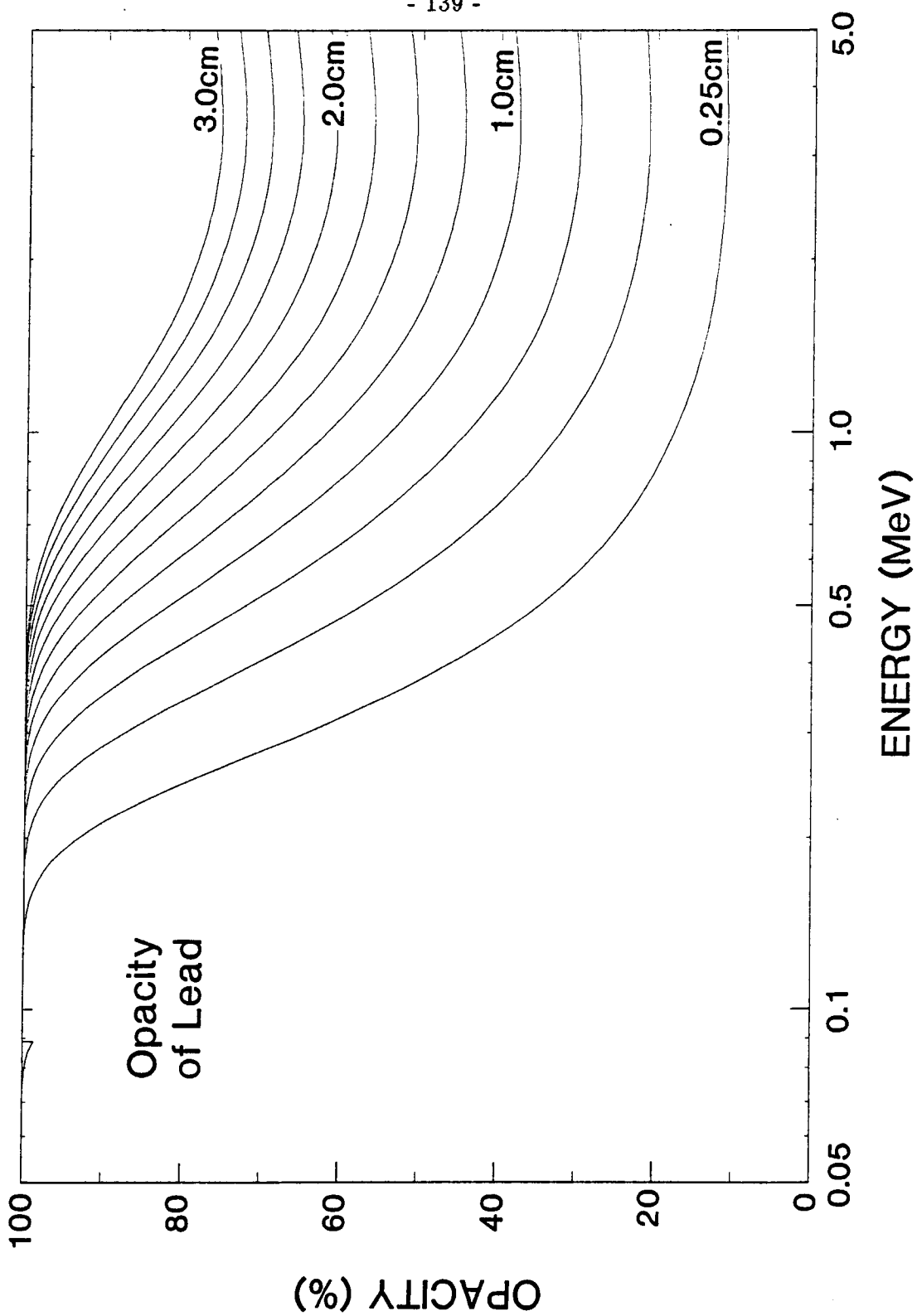
**Figure 4.3.** The GRIP mask. The mask used on GRIP is a 127 element URA. Each cell is 2.54 cm flat-to-flat, with the whole mask being 1.2 m in diameter. The black cells are cast lead hexagons 1.91 cm thick, supported on an aluminum honeycomb sandwich. During observations the mask rotates at a rate of one revolution per minute, allowing position-by-position background subtraction once every 20 seconds.



ORIGINAL PAGE IS  
OF POOR QUALITY



**Figure 4.4.** The opacity of lead versus energy for slabs of various thicknesses. The thickness changes between curves by 0.25 cm. The minimum opacity occurs near 3 MeV, where the pair-production and Compton-scattering cross-sections are comparable.



balloon instrument. A larger separation would require a larger mask to maintain the same field-of-view, as well as requiring more support structure.

The mask assembly is supported at the edges by a system of rollers on top of a conical tower attached to the shell. During observations the mask is rotated by a roller drive at a rate of one revolution per minute, providing position-by-position background subtraction once every 20 seconds.

The rotational orientation of the mask is encoded by two separate systems. The first system uses a linked chain that is embedded in the mask support ring. This chain is viewed by a combined LED and light detector that detects the passage of a link through its focal point. A reference is established by filling in a sequence of gaps between links. The chain has 1020 links providing  $\sim 0.1^\circ$  resolution. The second system uses a series of holes that have been drilled through the mask support ring. When a hole passes by an infrared LED the transmitted light is sensed by a small detector. The support ring has 24 evenly spaced holes, with an additional hole used to establish an absolute reference.

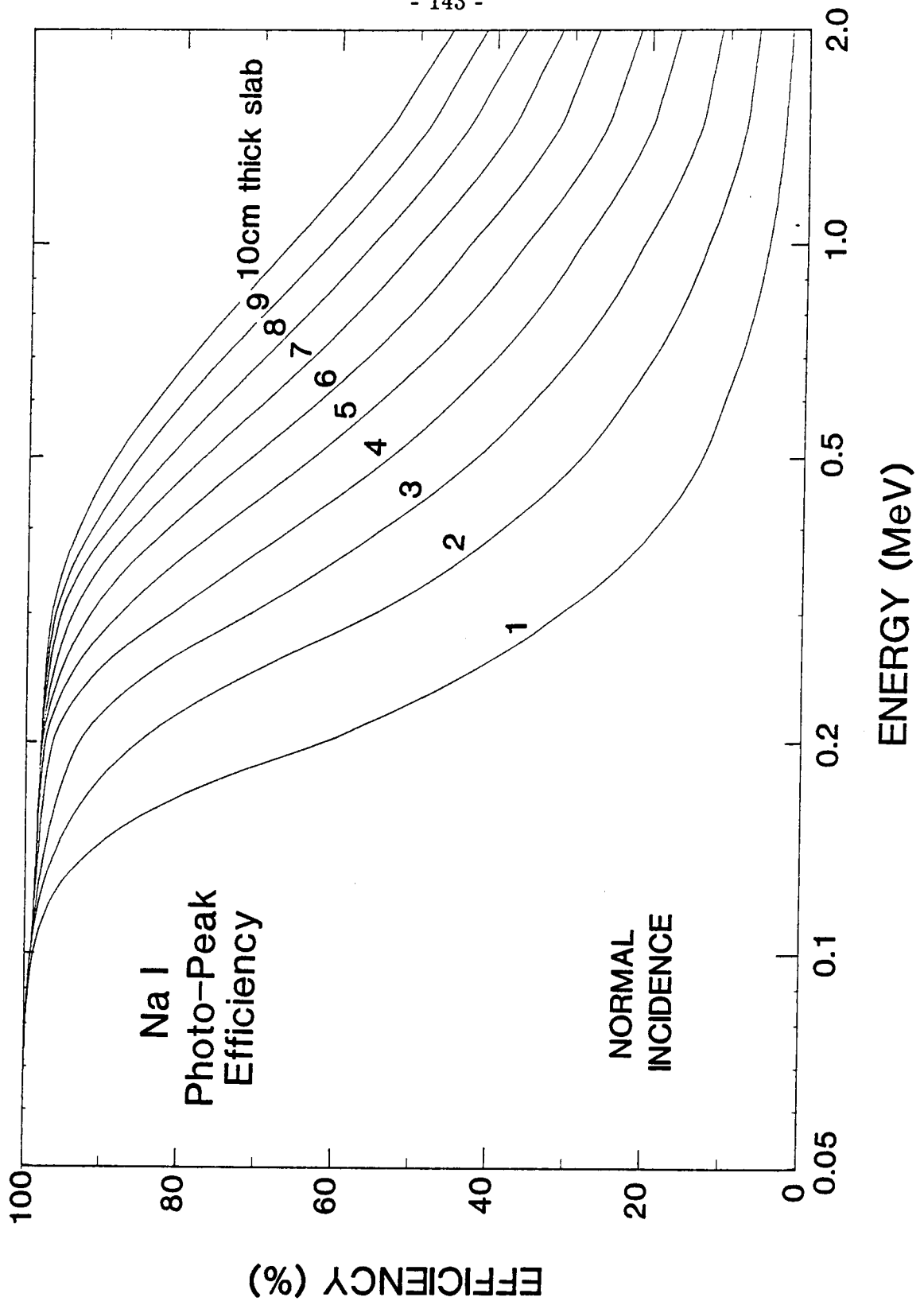
#### **4.2.2 The Primary NaI(Tl) Detector.**

The heart of the GRIP instrument is the primary NaI(Tl) detector. This is a NaI(Tl) scintillator instrumented in an Anger camera configuration. An Anger camera consists of a planar scintillator viewed by an array of PMTs, with the ratios of the PMT signals being used to determine the location of an interaction in the scintillator. Since its initial invention (Anger 1957), the NaI(Tl) Anger camera has been extensively developed for medical imaging at hard x-ray energies. The Anger camera was thus a natural and economical

choice as the position-sensitive detector of the GRIP gamma-ray telescope. A considerable effort was expended in the design and testing of the primary detector and its associated position algorithm. We will therefore begin this section by discussing the approach taken in the detector and position algorithm design. We will then describe the resulting detector and position algorithm and their coupled performance.

**Design Considerations.** Unfortunately much of the technology developed for medical imaging in the hard x-ray is not transferable to gamma-ray astronomy. The first reason for this is that the detectors used in the hard x-ray can be rather thin while gamma-ray detectors must be rather thick. Figure 4.5 shows how the photo-peak efficiency of a slab of NaI depends on its thickness. The curves, obtained by Monte-Carlo calculation, give the percentage of photons from a normally incident beam that deposit their full energy in the scintillator as a function of energy. At the lowest energies the photo-electric effect is the dominant form of interaction. Above 250 keV the majority of events suffer single or multiple Compton scatterings before any photo-electric interaction. At 100 keV, high efficiency can be achieved with a 1 cm thickness, however a 5 to 10 cm thickness is required at 1 MeV. Because of the thinness of an x-ray detector the PMT responses vary essentially two-dimensionally with respect to the photon interaction position. In contrast a thicker gamma-ray detector will have a more complex three-dimensional response. For thin x-ray Anger cameras, interaction positions can be successfully calculated by relatively simple algorithms based on the PMT position weighted sums of the PMT outputs. However, for a thicker gamma-ray camera, to successfully calculate the position of an interaction

**Figure 4.5.** The photo-peak efficiency of a slab of NaI. Each curve is labeled with the thickness of the NaI slab, and shows the percentage of photons from a normally incident beam that deposit their whole energy in the NaI. The curves are interpolated from Monte-Carlo calculations with 5000 photons per data point. The calculations neglect the escape of fluorescence x-rays. This surface effect, which is important below 70 keV, produces a decrease in photo-peak efficiency which is nearly independent of slab thickness.



within the plane of the detector, we must also calculate the depth of the interaction within the detector.

A second reason that the hard x-ray medical technology is not easily transferred to gamma-ray astronomy involves the relationship between the design of the detector and the design of its position algorithm. A medical imaging system is required to perform real time calculations of event locations at high event rates for x-ray interactions at well-known energies. To achieve fast calculations, simple position algorithms that may be performed with analog electronics are employed. The detector geometry is then tailored to optimize the performance of the chosen position algorithm. This often involves reducing distortion or enhancing position resolution at the expense of energy resolution. For a gamma-ray astronomy detector however, we wish both good spatial and energy resolution, and the complexity of the position algorithm is, within limits, secondary. We therefore decided that in evaluating the design of a detector we would assume that the "best possible" position algorithm was to be used. We took as a working definition of the "best possible" position algorithm a maximum likelihood calculation of the event positions.

**Maximum Likelihood Positions and Energies.** The maximum likelihood method requires a detailed knowledge of the characteristics of a detector. For each PMT we need to know the mean number  $N_i(\vec{x}, E)$  of photoelectrons expected to be collected for an interaction at position  $\vec{x}$  within the detector depositing energy  $E$ . For multiple-interaction events we would ideally like to determine the position of the first interaction. We will however assume that we can only determine the energy weighted mean position of the total interaction. We will further assume that the expected PMT responses for a



multiple-interaction event are identical to the responses expected for an event which deposits its total energy at the mean interaction position.

The maximum likelihood position  $\bar{x}$  and energy  $E$  are found by maximizing the likelihood function, or equivalently, its logarithm. Since the statistical variation of the number  $n_i$  of photo-electrons produced in PMT  $i$  is governed by the Poisson distribution, the logarithm of the likelihood function is given by:

$$\ln L(\bar{x}, E) = \sum_i [n_i \ln N_i(\bar{x}, E) - N_i(\bar{x}, E)] + \text{const.} , \quad (4.1)$$

where  $N_i(\bar{x}, E)$  is the response function for PMT  $i$  and gives the expected number of photo-electrons as a function of interaction location and gamma-ray energy, and the constant depends only on the  $n_i$ .

The problem simplifies if we assume that the response functions are separable functions of  $\bar{x}$  and  $E$ :

$$N_i(\bar{x}, E) = l(E) f_i(\bar{x}). \quad (4.2)$$

The PMT responses are approximately proportional to gamma-ray energy at a given  $\bar{x}$ . Small deviations from proportionality are due to the non-linearity of the NaI(Tl) light response and are accounted for by the function  $l(E)$ . We will choose the normalization of  $l(E)$  so that it represents the average number of photo-electrons collected if all of the photons produced are absorbed by a photo-cathode.

We may solve  $\partial \ln L / \partial E = 0$  to obtain an implicit expression for  $E$ :

$$l(E) = \frac{n_{tot}}{C(\bar{x})} \quad (4.3)$$

where  $n_{tot}$  is the total number of collected photo-electrons

$$n_{tot} = \sum_i n_i \quad (4.4)$$

and  $C(\vec{x})$  is the light collection efficiency given by

$$C(\vec{x}) = \sum_i f_i(\vec{x}) . \quad (4.5)$$

Upon substitution of (4.2) and (4.3) into equation (4.1) we find that  $\ln L$  reduces within constant terms to:

$$\ln L(\vec{x}) = n_{tot} \sum_i \hat{n}_i \ln \hat{N}_i(\vec{x}) \quad (4.6)$$

where the PMT responses  $\hat{n}_i$  and the response functions  $\hat{N}_i$  are now normalized:

$$\hat{n}_i = \frac{n_i}{n_{tot}} \quad (4.7)$$

and

$$\hat{N}_i(\vec{x}) = \frac{f_i(\vec{x})}{C(\vec{x})} . \quad (4.8)$$

From equation 4.6 we see that the maximum likelihood solution for the interaction position is independent of the energy of the interaction and the total number of photo-electrons collected, depending only on the normalized PMT signals  $\hat{n}_i$ . The information needed for the calculation of interaction positions consists of the normalized response functions  $\hat{N}_i(\vec{x})$ .

The covariance matrix  $\sigma_{ij}$  associated with the statistical position errors for an interaction at position  $\vec{x}$  may be determined by evaluating the second derivatives of the likelihood function at  $n_i = l(E)f_i(\vec{x})$ :

$$(\sigma)_{jk}^{-1} = l(E)C(\vec{x}) \sum_i \frac{1}{\hat{N}_i(\vec{x})} \frac{\partial \hat{N}_i(\vec{x})}{\partial x_j} \frac{\partial \hat{N}_i(\vec{x})}{\partial x_k} . \quad (4.9)$$

The inverse of the covariance matrix is often referred to as the "information matrix." As can be seen, the most information about the interaction occurs when a PMT with a small response has a large derivative. If at any position, such as near the edge of the detector, only two PMTs have rapidly varying responses, then there will be an axis along which there is little information about the interaction position. The normalization of the responses shows that the maximum likelihood method results in a position algorithm that is independent of energy. Unfortunately this also obscures two important features. First, gradients in the light collection efficiency introduce correlations between the determined position and energy. Second, the statistical error associated with determining the total light emitted is reduced by improving the light collection efficiency. It should be noted however that there is another source of error in determination of the interaction energy other than the photo-electron statistics. This is due to the nonlinear response of NaI at low energies which for multiple Compton scatterings causes the total light output to depend on the distribution of the energy deposited among the individual interactions.

**Design Evaluation.** With a model of the PMT responses equation (4.9) gives a method for evaluating the photo-electron statistics component of the

position resolution of a particular detector. We used this method to compare several detector designs. The detectors considered consisted of a cylindrical crystal of NaI(Tl) viewed through a glass window by an array of PMTs. The detector was assumed to be enclosed in a housing coated with a diffuse reflector. The number, size and spacing of the PMTs as well as the thickness of the detector and various surface characteristics varied with each design. For each design the PMT responses were determined by Monte-Carlo simulation. The simulations involved the propagation of photons from a point source in the detector until they were either absorbed or produced a photoelectron in one of the PMTs. The simulations included detailed models of the reflective and absorption properties of photo-cathodes and the scattering at the rough surfaces of the NaI detector.

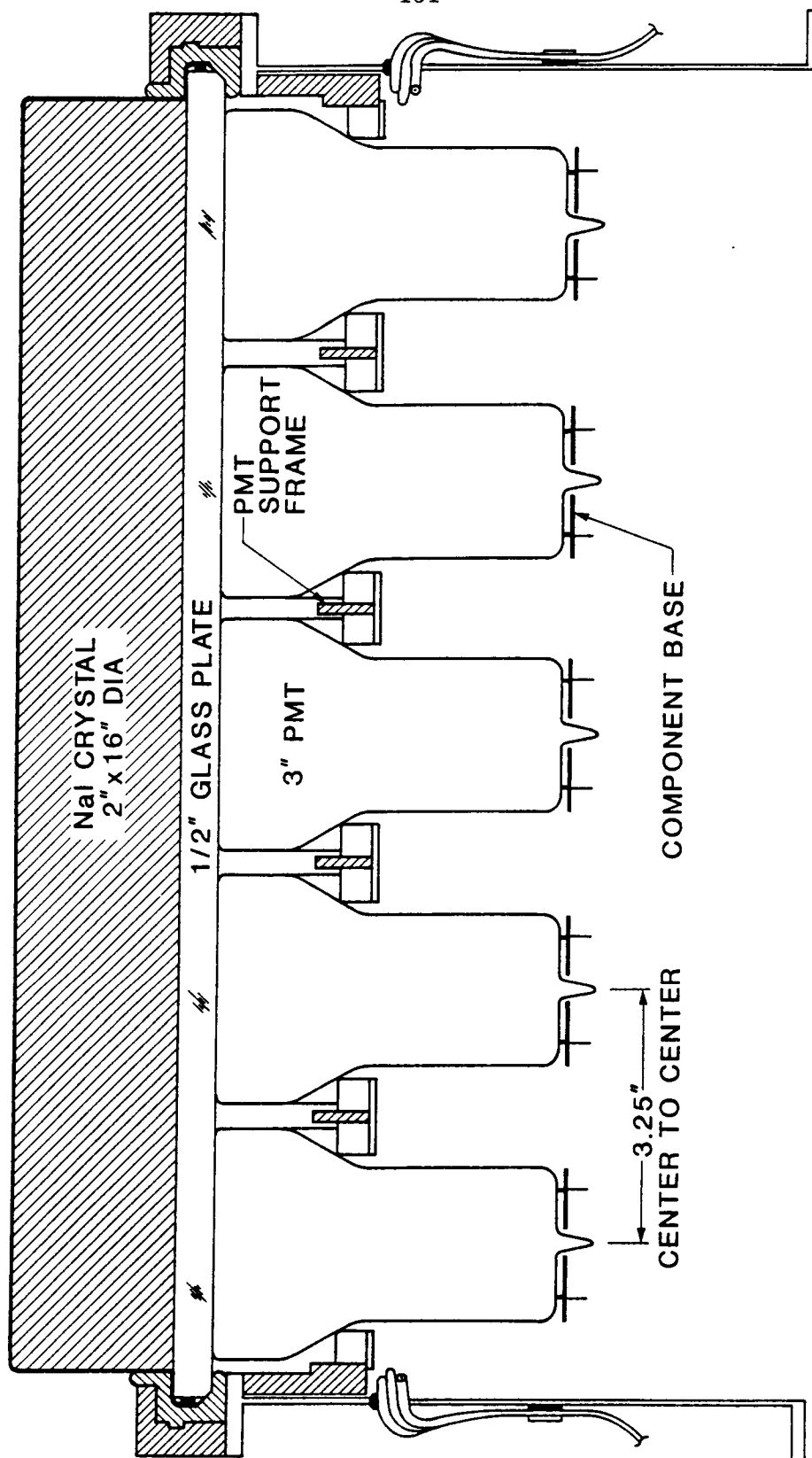
Equation (4.9) requires the calculation of the derivatives of PMT responses, which are approximated by finite differences. Fortunately there is a simple technique for obtaining Monte-Carlo calculations of differences that greatly reduces the computation time needed for a result of given accuracy. To find the difference in the response of a given PMT between interaction positions A and B we propagate pairs of photons, one starting at position A and one at position B. Both photons in the pair are required to use the same series of random numbers, so that their fate would be identical if they started from the same position. We gather as statistics  $n_A$  the number of times the PMT collected the photon from A but not from B,  $n_B$  the number of times the PMT collected the photon from B but not from A, and  $N$  the total number of pairs generated. Then the estimate of the response difference is  $(n_A - n_B)/N$ . The number of times that the PMT collected both photons does

not enter into calculation of the response difference, and therefore does not contribute to its statistical error. This procedure may be justified and generalized by viewing the Monte-Carlo calculation as a volume integration in the space of random number sequences.

By Monte-Carlo evaluation of (4.9) at several points for candidate designs, we reached several conclusions. The responses of all thick detectors showed marked dependence upon the depth of interaction in the detector. Thus a thick camera plate can provide depth information as well as positions in the plane of the detector. In general as much of the surface of the detector as possible should be covered with PMTs. Decreasing the PMT diameter beyond the thickness of the detector seems to have little effect on the resolution attainable in the plane of the detector, with this resolution scaling with the detector thickness. It should be noted that the resolution derived via equation (4.9) is a purely local measurement. In particular, it is quite possible that two separated locations have very similar responses, and would therefore not be distinguishable.

**The Detector and Instrumentation.** A cross-section of the detector assembly designed for GRIP is shown in Figure 4.6. The 41 cm diameter, 5 cm thick NaI(Tl) camera plate is a "polyscin" forging which has been rough polished and bonded to a 1.27 cm glass optical window. The hygroscopic NaI is protected from the atmosphere by a thin ( $\sim 0.8$ mm) Al housing that is hermetically bonded to the optical window. The air gap between this housing and the NaI insulates the crystal from thermal shock. The inside of this housing is coated with a diffusely reflective white material to improve light collection efficiency. The optical window is connected to an Al support ring by

**Figure 4.6.** The Primary Detector Assembly. The primary NaI(Tl) detector is shown with its housing and PMTs. The PMTs are attached with silicon grease and supported by a phenolic plate and a system of rubber washers.



RTV potting which protects the NaI from mechanical shock. The NaI camera plate with window, hermetic seal, and potting was manufactured by the Harshaw Chemical Co. This assembly is mounted in an Al housing that provides mechanical support and a light seal.

Nineteen 3-inch Hamamatsu R1307 photomultiplier tubes view the NaI through the the optical window. The PMTs are optically coupled to the glass with Dow Corning Q2-3067 optical coupling compound. The PMTs are confined in a hexagonal close-packed array with a spacing of 8.25 cm by a phenolic plate and a system of rubber washers. The area of the optical window not covered by PMTs was covered by white cardboard to enhance light collection efficiency.

The PMTs are biased with negative high voltage applied to the photocathodes. This negative biasing was chosen to allow a direct coupling of the anodes to their pre-amplifiers. To avoid heat sources near the NaI, seven voltage levels were generated externally and distributed to the PMTs on a PC board. The last six dynodes of each PMT are connected to the voltages between -100V and -600V in 100 volt steps through 50k $\Omega$  resistors. The resistors limit the current to the PMTs and allow prompt recovery from the passage of charged particles through the detector. The photocathode, grid and first two dynodes are biased through a resistive divider placed between -1000v and -600v. This divider contains a resistor which alters the total voltage drop between the photocathode and the third dynode, allowing the gain of each PMT to be adjusted. Before installation the gains of the PMTs were balanced.

Each PMT in the primary detector has associated with it amplification, integration and analog to digital conversion electronics. These electronics are



located at the bottom of the pressure vessel. Each PMT anode signal is pre-amplified and passed to a gated integrator. When the event logic, which is discussed below, determines an event should be processed, it begins by initiating the integration of the pre-amplified PMT signals. This integration continues for 1.5  $\mu$ s after which time each integrated signal is sampled and held. The held signals are then converted from an analog to digital format. The analog to digital converters (A/Ds) produce a 12 bit output. The gain of the signal chain is set so that an energy deposit of  $\sim 2.5$  MeV immediately above a PMT will produce a full scale output.

From the sum of the pre-amplified PMT signals, three discriminator signals are derived. These discriminators are the Zero Level, Lower Level and Upper Level Discriminators (ZLD, LLD, and ULD), which are used by the event logic. The discrimination levels of the ZLD, LLD and ULD are all commandable. The ZLD has the lowest discrimination level and is used to detect an energy deposit and is the basis for timing the processing of an event. Events with energy deposits outside a given range may be rejected for processing by the use of the LLD and ULD.

**Detector Response Measurements.** Use of the maximum likelihood method to determine interaction positions requires mapping of the responses of the PMTs. We obtained knowledge of the required response functions using the 662 keV line of a collimated  $^{137}\text{Cs}$  source. The gamma-ray beam was directed at normal incidence to the camera plate and could be positioned laterally to an accuracy of 0.1 mm by an x-y translation stage. The beam spread at the center of the NaI(Tl) crystal was 3 mm FWHM. Data were obtained for source locations chosen to lie on a hexagonal grid of spacing 2.38

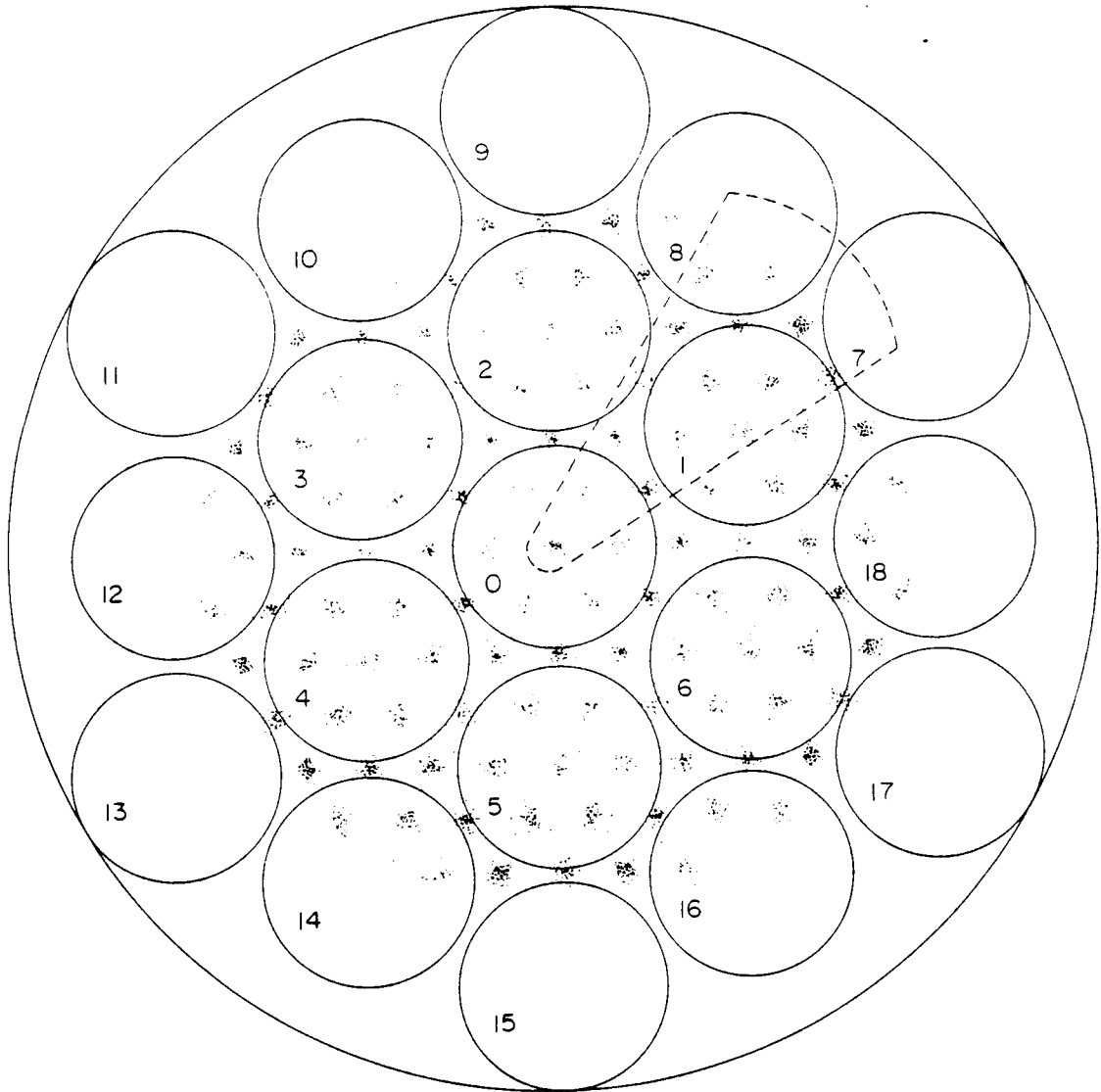
cm and covering a wedge-shaped region, shown in Figure 4.7, of approximately  $1/12$  of the area of the camera plate. Data from additional source locations were taken to verify the symmetry of the camera plate response.

For the response mapping the camera plate PMTs were instrumented to closely approximate the desired situation in which each PMT is individually pulse-height analyzed for each gamma-ray event. The only deviation from such a scheme was required by the 15 input limitation of the pulse-height analysis system available at that time. Twelve of the nineteen PMTs (0 through 10 plus 18, as labeled in Figure 4.7) were individually pulse-height analyzed, while the remaining three input channels were used to analyze three sum signals: (11+12), (13+14+15), and (16+17). We then confined the majority of our measurements to the wedge-shaped region shown in Figure 4.7 so that the summed signals would be small and contain little of the relevant information on gamma-ray interaction location. For each gamma-ray event all 15 channels were pulse-height analyzed and the resulting data were stored on magnetic tape for subsequent computer study. The GRIP instrument employs a nineteen input analysis system and the algorithm devised from the initial measurement have been tested with its use.

The data for each gamma-ray beam location were analyzed separately. After making preliminary cuts to select only photopeak events and to reject background events occurring outside the beam, the data for each beam were fit to a straight line in the space of PMT signals.

**Figure 4.7.** The detector response measurement grid and PMT numbering system. The figure shows the positioning of the PMTs on the NaI detector and the number system referred to in the text. Response measurements were made on the grid of points shown, primarily within the outlined wedge.

ORIGINAL PAGE IS  
OF POOR QUALITY



The initial cuts were made using a simple linear position algorithm to eliminate events far from the beam location. Inspection and understanding of the remaining data were difficult because of the 15-dimensional nature of the data space. We knew however that events near any location in that space should be locally confined to three dimensions corresponding to the three spatial dimensions. We therefore employed a technique that selected those linear combinations of normalized PMT signals that showed the statistically largest variations in the remaining data set.

This procedure began by finding the means  $\mu_i$  of the normalized PMT signals  $\hat{n}_i$  for the single beam data set. Then new variables were formed that near the mean had equal errors and were statistically independent (except for the normalization):

$$a_i = \frac{\hat{n}_i - \mu_i}{\hat{n}_i} \quad . \quad (4.10)$$

A new basis  $\bar{e}^j$  was then found for the data vectors  $\bar{a}$ . In this basis the correlation matrix

$$V_{ik} = \frac{1}{N} \sum_{events} a_i a_k \quad (4.11)$$

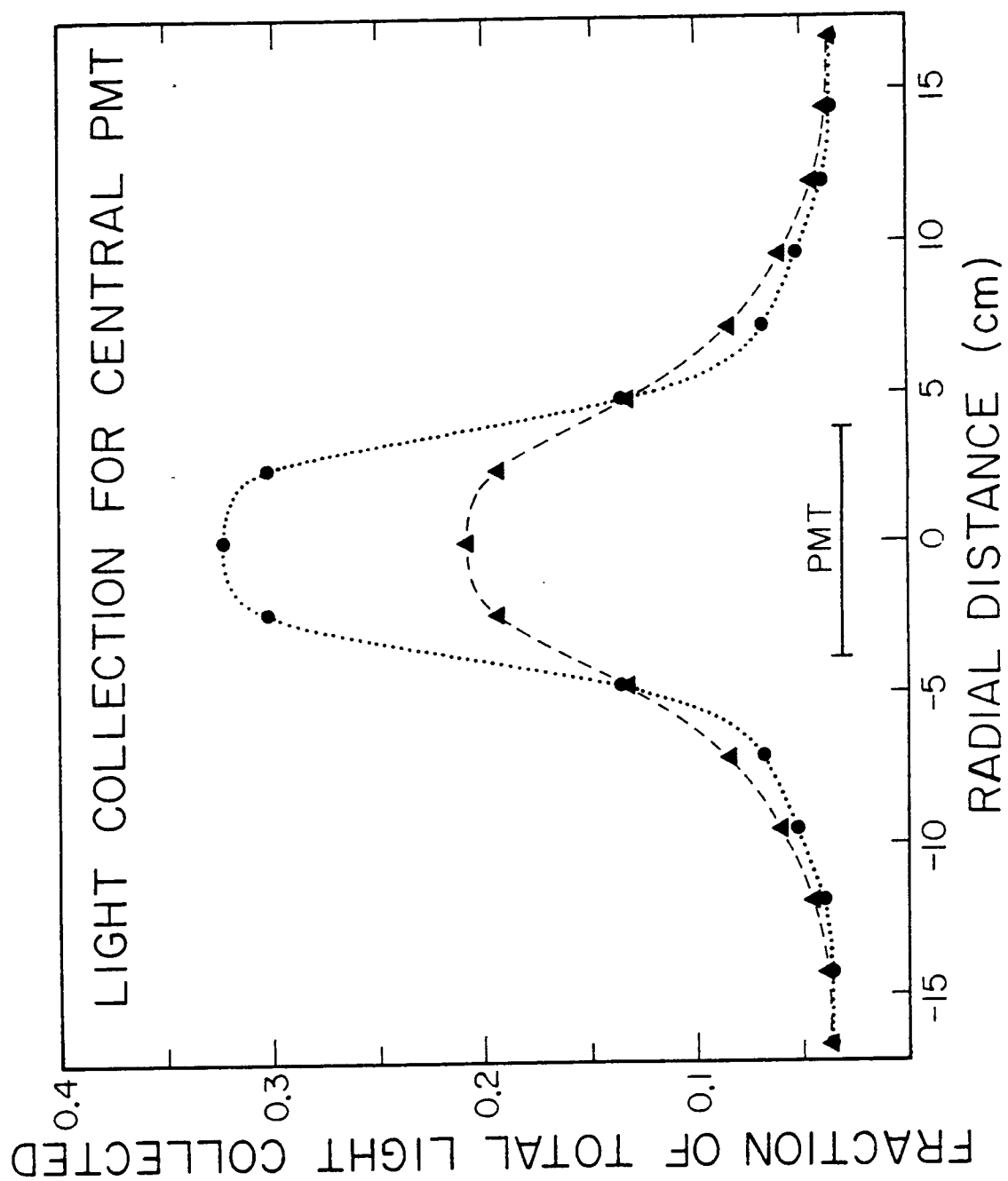
is diagonalized. Here  $N$  is the number of events. The eigenvalues of the correlation matrix determine the statistical significance of the variation along each axis  $\bar{e}^j$ . In general the data sets showed only three or four dimensions with significant variation. The data were then examined only in this smaller subspace, and closer cuts were made to reject background events outside of the beam.

The PMT responses are in general non-linear, however the responses within a single beam appeared to lie along a line in the data space. This was found to be a fair approximation for all the beam data sets. The data were therefore fit to lines in the data space with the parameter  $\lambda$  of distance along the line defined such that  $\lambda = 0$  corresponded to gamma-ray interactions near the front of the crystal, while  $\lambda = 1$  corresponded to events at the back of the crystal, nearest the PMTs.

Thus the normalized response functions determined from the beam measurements were expressed as functions of a depth parameter  $\lambda$ , rather than the physical depth  $z$ . We found that the depth parameter  $\lambda$  is not a strictly linear function of  $z$  and that the relation between  $\lambda$  and  $z$  varies as a function of the lateral  $(x,y)$  position. However, since our primary interest is in an optimum algorithm for the lateral position  $(x,y)$  we substituted  $\lambda$  for  $z$  in the maximum likelihood formulation and below we will refer to the event location  $\vec{x} = (x,y,\lambda)$ .

In general the PMT responses were found to have a marked dependence on the depth of a gamma-ray interaction. This depth dependence may be seen in Figure 4.8 which shows the response of the central PMT as deduced from the 662 keV beam measurements. When the collimated beam was directed at normal incidence to the camera plate and positioned opposite the central PMT, the fraction of the total light signal collected by the central PMT increased with the depth of gamma-ray penetration. The signal was a maximum of 33% for gamma-ray interactions occurring deep in the camera plate and close to the PMT face, but was only 22% for interactions near the front surface of the camera plate. As the beam is moved toward the edge of

**Figure 4.8.** The response of the central PMT. The figure gives the ratio of the light collected by the central PMT to the total light collected versus the distance of the interaction from the center of the detector. The solid circles interpolated by a dotted line are the response at the back of the detector near the PMTs. The triangles interpolated by a dashed line represent the response at the front of the detector. Data at negative distances are obtained by symmetry.





the detector the response falls off, reaching a minimum value of 3.5% independent of the interaction depth. The width of the response function corresponds roughly with the spacing of the PMTs, and is largest toward the front of the detector.

**The Position Algorithm.** While the details of the method by which the likelihood function is maximized are not critical to the results, we include for completeness a discussion of the particular approach we used.

The maximization of the likelihood function can be done by an iterative technique where a first guess  $\vec{x}^0$  of the position is assumed to be close to the position  $\vec{x} = \vec{x}^0 + \Delta\vec{x}$  of the maximum. This first guess can be found for example by determining which PMT has the largest signal. The displacement  $\Delta\vec{x}$  is estimated by setting

$$\frac{\partial \ln(L(\vec{x}^0 + \Delta\vec{x}))}{\partial x_k} \approx \frac{\partial \ln(L(\vec{x}^0))}{\partial x_k} + \sum_l \frac{\partial^2 \ln(L(\vec{x}^0))}{\partial x_k \partial x_l} \Delta x_l = 0. \quad (4.12)$$

The solution for  $\Delta x_l$  is then used to compute a refined guess of the position, and the procedure iterated.

Our algorithm approximates this procedure in two ways. First, the positions  $\vec{x}^0$  are restricted to the lattice of points at which we have measured the PMT responses (or inferred the responses by symmetry). Second, we solve equation (4.12) for  $\Delta x_l$  with the assumption that deviations of the PMT responses  $\hat{n}_i$  from the mean responses  $\hat{N}_i(\vec{x}^0)$  are small. The resulting solution is:

$$\Delta x_l = \sum_i X_{li} (\hat{n}_i - \hat{N}_i(\vec{x}^0)) \quad (4.13)$$

where

$$X_{li} = \sum_k \eta_{lk} \frac{\partial \ln(\hat{N}_i)}{\partial x_k} \quad (4.14)$$

and

$$(\eta^{-1})_{lk} = \sum_i \frac{1}{\hat{N}_i} \frac{\partial \hat{N}_i}{\partial x_l} \frac{\partial \hat{N}_i}{\partial x_k} \quad (4.15)$$

with all evaluations performed at  $\bar{x}^0$ .

The required derivatives of the normalized response functions  $\hat{N}_i$  were computed by finite differences on the lattice, and for each lattice point we stored  $\hat{N}_i$  and  $X_{li}$ . Then equation (4.13) defines a linear algorithm valid for calculating the event position near  $\bar{x}^0$ . At a moderate distance ( $\sim 2-3\text{cm}$ ) from this point, distortions appeared as a compression or expansion of the x and y axes with increasing depth. This is unavoidable for any purely linear algorithm, although the magnitude of the distortion could be reduced by working on a finer lattice. To correct for this distortion we introduced a quadratic correction :

$$\Delta x_l' = \Delta x_l + \sum_{mn} C_{lmn} \Delta x_m \Delta x_n \quad (4.16)$$

where

$$C_{lmn} = -\frac{1}{2} \sum_i X_{li} \frac{\partial^2 \hat{N}_i}{\partial x_m \partial x_n} . \quad (4.17)$$

This correction is applied only after iteration of equation (4.13) has converged.

**Detector and Position Algorithm Performance.** In evaluating the success of the detector and position algorithm design we will consider systematic distortions in positions, position resolution, and energy resolution.

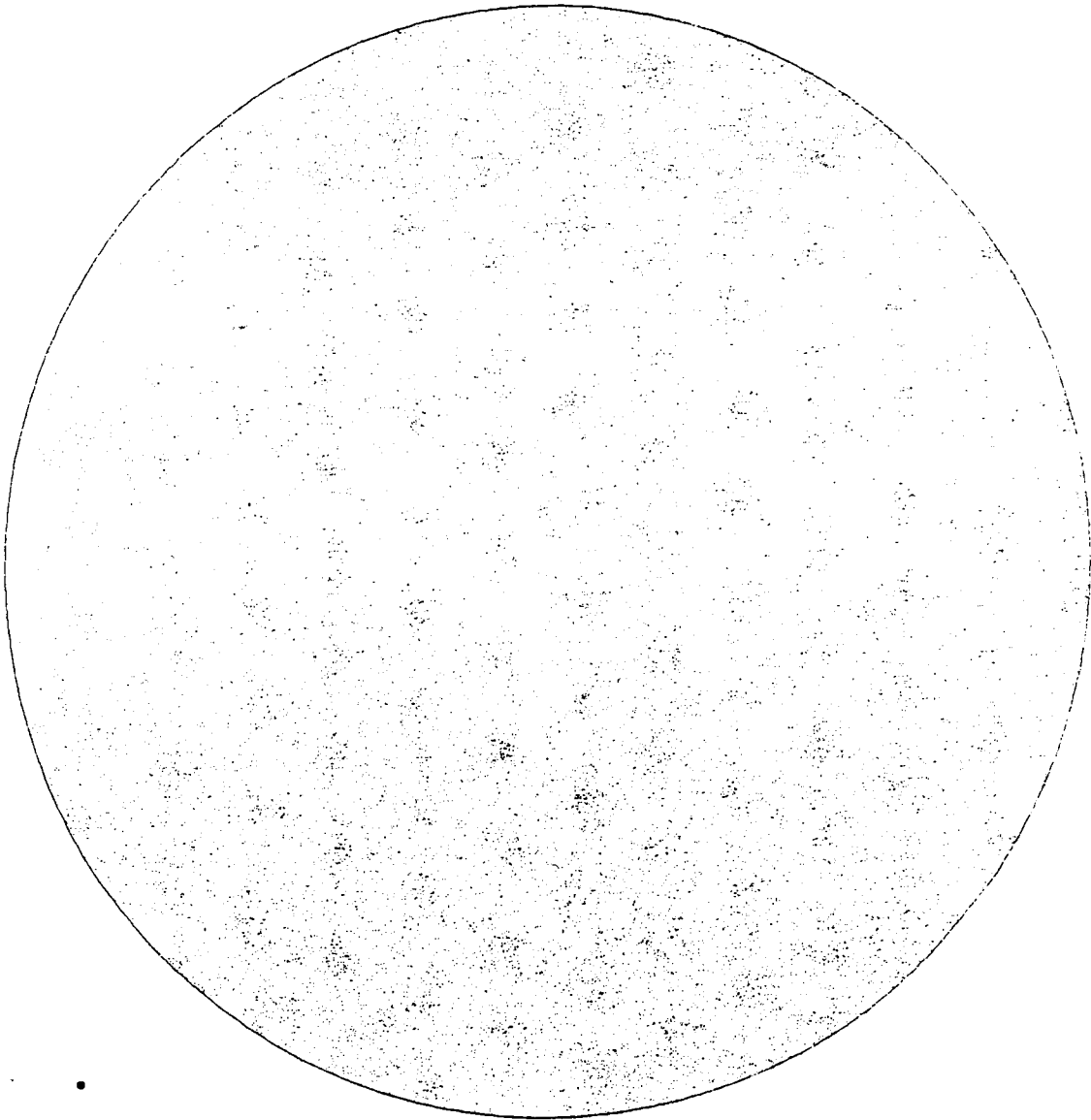
For preflight calibration of the detector, a 0.5 inch thick lead plate with an array of 0.25 inch lead holes is placed in front of the primary NaI(Tl) detector. The detector is then exposed through the plate to a gamma-ray source, so that in effect the detector is exposed to an array of gamma-ray beams of known location. The holes on the plate are arranged in a hexagonal array with 2.38 cm spacing. The plate may be accurately located in three positions with respect to the detector, allowing the measurement of the detector response to a composite hexagonal array of beams with 1.37 cm spacing.

Figure 4.9 shows calculated positions of gamma-rays from a  $^{133}\text{Ba}$  source, for an off-center position of the lead hole plate. The events have been restricted to the 356 keV line of the  $^{133}\text{Ba}$  source, and are displayed only within a 28 cm diameter region centered on the detector. This region is considered the usable area of the detector, with the remainder being employed as an active shield. Due to the collimation of the source, data is less dense at the left edge of the plot. A comparison of the reconstructed beam positions with their expected locations has revealed no global distortions. There are however systematic errors in the beam locations that for a few points are as large as 2 mm. These errors are due to inaccuracies in the measurement of the PMT responses. These systematic position errors affect an image by broadening the point source response function and introducing errors in the peak position that are smaller than 2 arc minutes. Further analysis of the calibration data will allow improved measurement of the response functions, which

**Figure 4.9.** Hole plate calibration data. The points are calculated positions of events obtained during a hole plate calibration with a  $^{133}\text{Ba}$  source. Events have been restricted in energy to the 356 keV line. The hole plate grid, which has 2.38 cm hole spacing, has been placed in an off-center location. The circular region is at a radius of 14 cm.

ORIGINAL PAGE IS  
OF POOR QUALITY

- 165 -



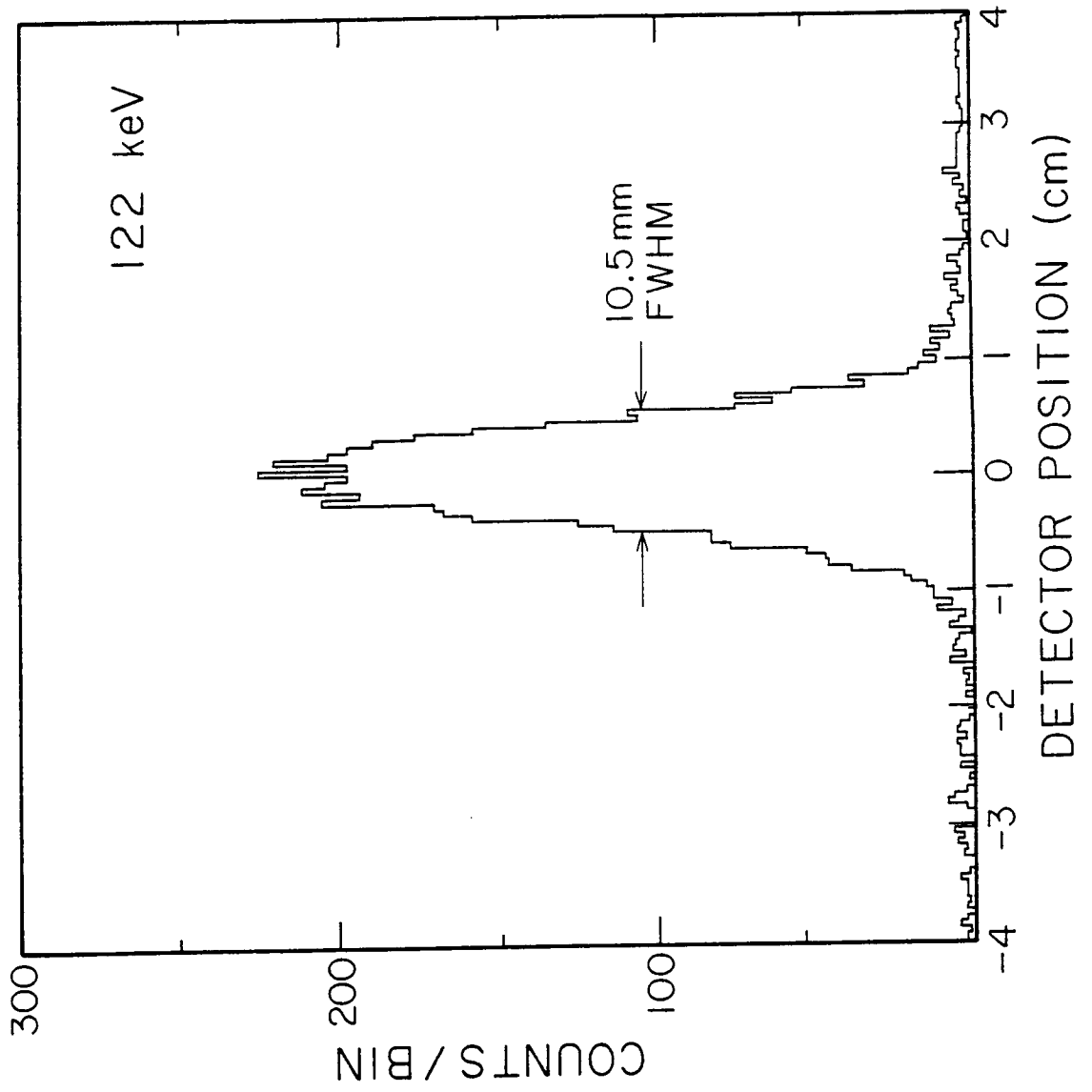
should reduce the systematic errors in the position reconstruction.

The position resolution obtained at the center of the camera plate is illustrated in Figures 4.10 and 4.11 at gamma-ray energies of 122 keV and 662 keV respectively. This data was obtained with collimated sources when the PMT responses were being measured on a x-y translation stage. The calibration plate data is unsuitable for measuring the detector's position resolution because of the large hole size. Since at 122 keV the position broadening is due almost entirely to photo-electron statistics, the distribution of measured positions seen in Figure 4.10 is approximately Gaussian. In contrast, the distribution of measured positions obtained at 662 keV is distinctly non-Gaussian, with extended tails due to Compton scattering.

Figure 4.12 shows a calculation of the expected position resolution as a function of energy. The calculation employed a Monte Carlo gamma-ray propagation code to determine the contribution to the FWHM position resolution due to Compton scattering which becomes important above an energy of approximately 300 keV. The Gaussian contribution, due to photo-electron statistics and light collection effects, was normalized to the measured position resolution at 122 keV where the effect of Compton scattering is negligible. The calculated curve agrees well with the data point at 662 keV which has been adjusted down from 7.0 to 6.3 mm to account for the collimated beam width of 3.0 mm FWHM. Figure 4.12 indicates that in spite of Compton scattering the expected FWHM position resolution continues to improve up to the limit of the calculation at an energy of 2 MeV.

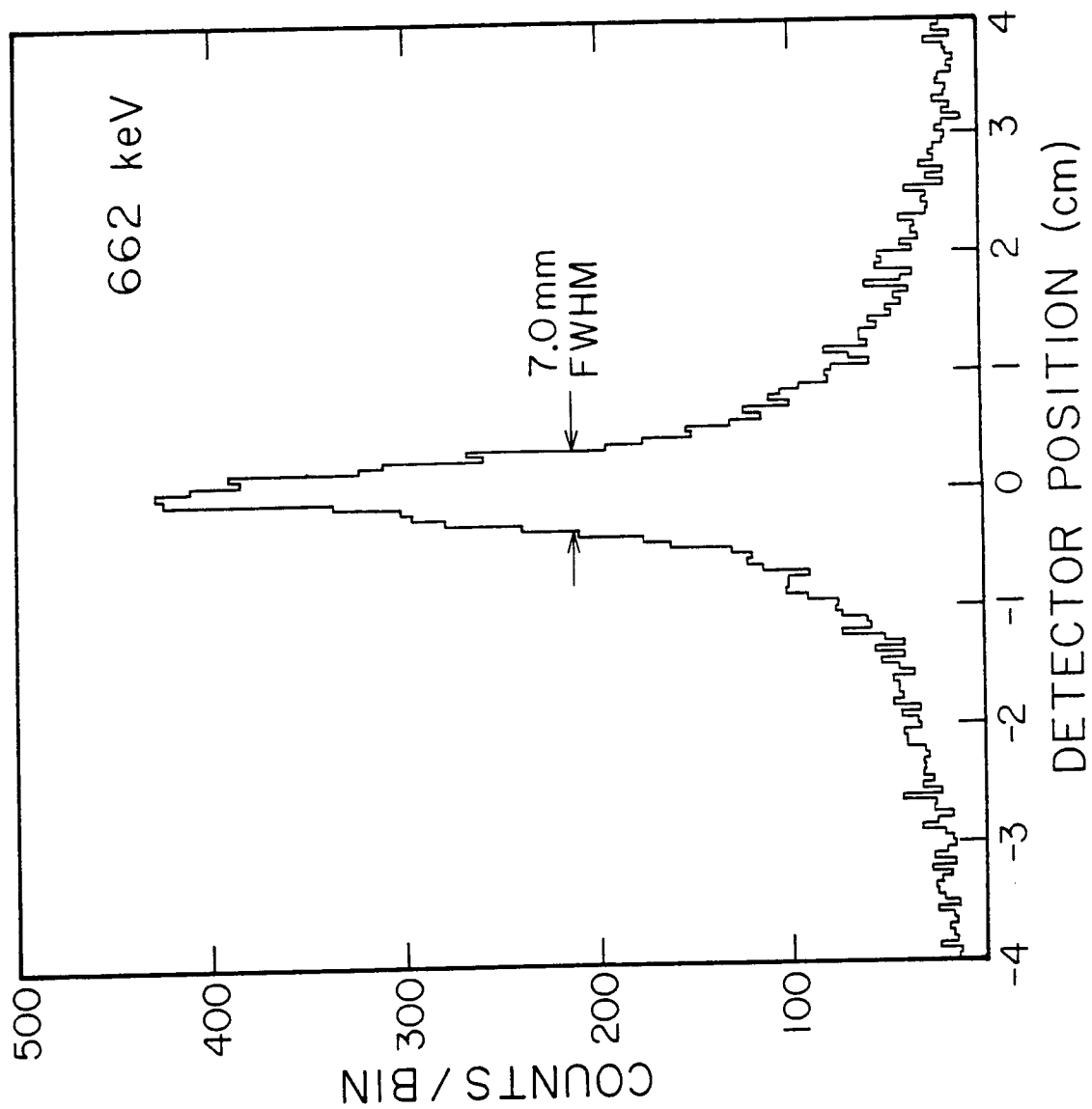
Towards the edge of the detector the position resolution is degraded. This degradation is to be expected because of the light reflected off the NaI

**Figure 4.10.** Position Resolution at 122 keV. The figure shows the measured point spread function at the center of the detector at 122 keV. The width of the peak is primarily due to photo-electron statistics.

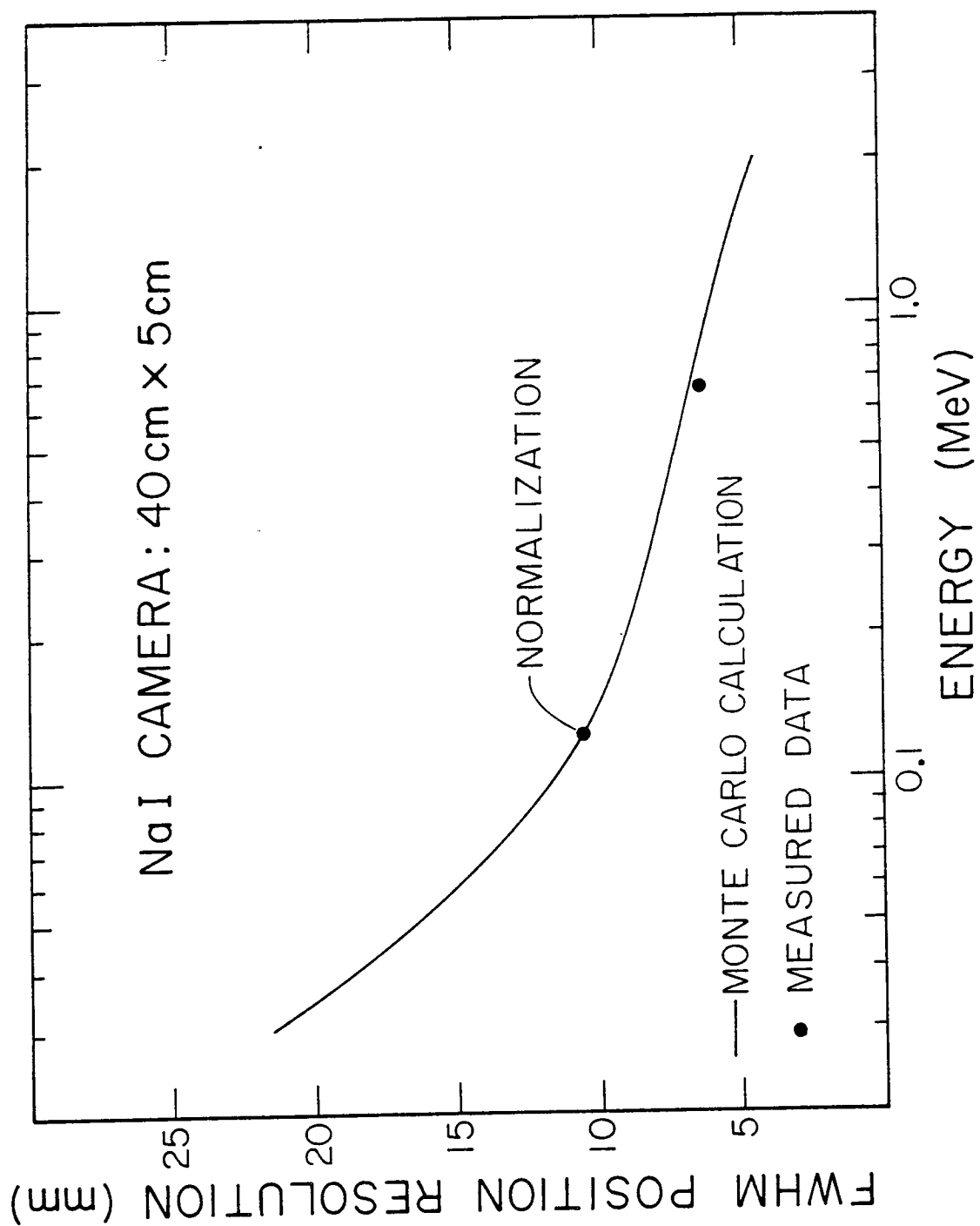




**Figure 4.11.** Position resolution at 662 keV. The figure shows the measured point spread function at the center of the detector at 662 keV. The wide tails are due to Compton scattering.



**Figure 4.12.** Calculated position resolution versus energy. The figure gives the expected detector position resolution based on energy scaling of the photo-electron contribution to the point spread function, and Monte-Carlo calculation of the Compton scattering contribution to the point spread function. The photon-statistical contribution is normalized to the measurement at 122 keV. The FWHM of the point spread function continues to decrease with energy, with the main effect of Compton scattering being in the wings of the distribution.



crystal's sides cannot be distinguished from that coming directly from the event. Thus while axial resolution is maintained, there is little information to separate the radial and depth dependence of the signals. The degradation is further compounded by the incomplete PMT coverage near the detector's edge. We have chosen a region within 14 cm of the center of the detector as usable for imaging. In retrospect, this region could have been enlarged if additional PMTs were used to cover the edge of the detectors face. Instrumentation of the sides of the NaI crystal could possibly make the whole detector usable.

The light collection efficiency of the detector varies by 10% over the usable area of the detector. A large component of this variation is a simple radial falloff. At present a complete measurement of the light collection efficiency has not been made. However, when a simple radius-squared correction is made for the light collection efficiency, the usable area of the detector has a energy resolution of 9% FWHM at 662 keV. This is to be compared to the 7% FWHM at 662 keV obtained for a beam at the center of the detector.

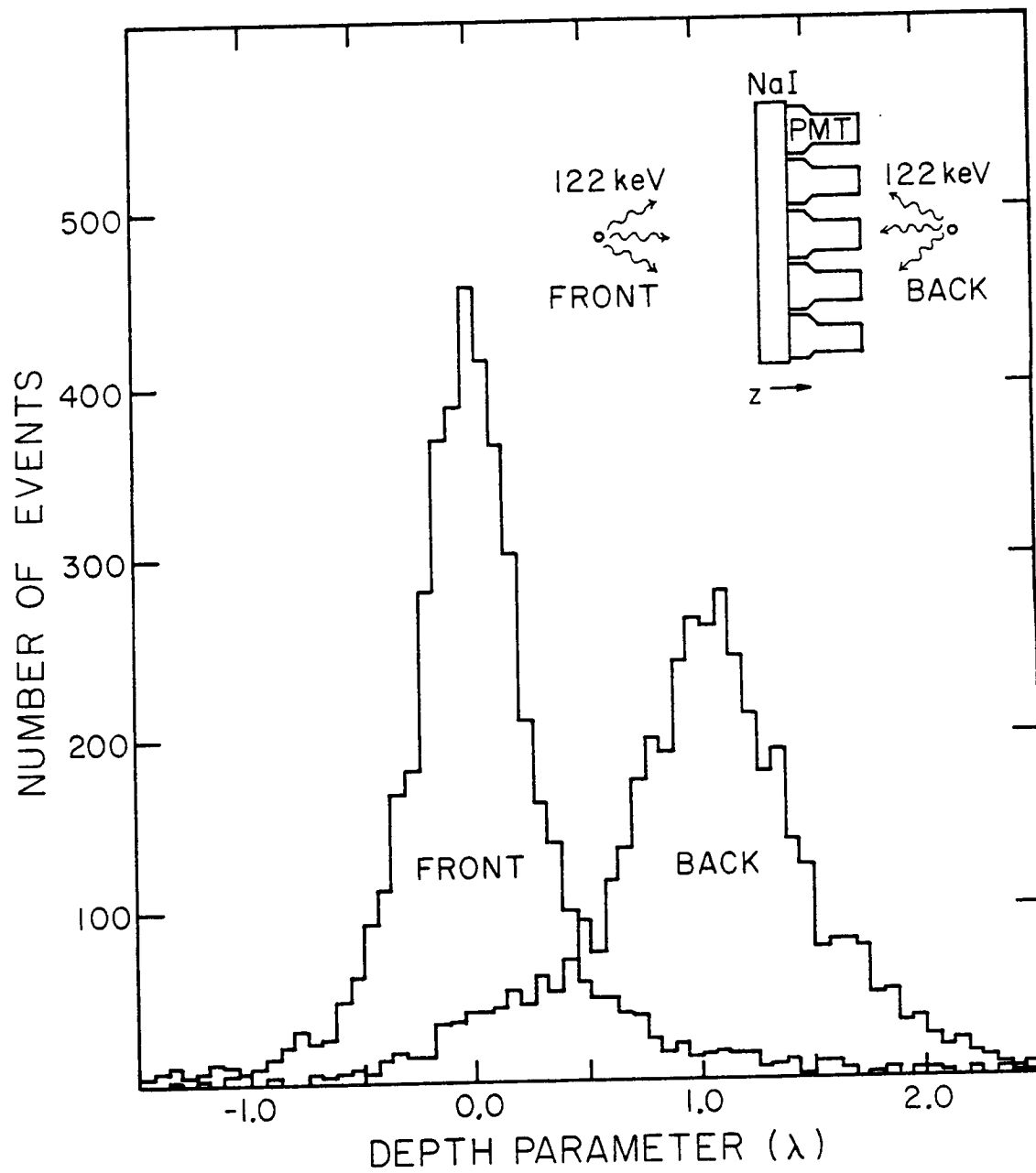
A consequence of the development of a three-dimensional position-determination algorithm is the capability of measuring the depth of the gamma-ray interaction in the detector. This depth determination capability is potentially very useful since it allows a reduction of background. For example, low-energy gamma-rays are expected to interact primarily near the front face of the crystal. The interaction length for gamma-rays of less than 300 keV energy is less than 1.5 cm. Thus, low-energy events with interaction positions in the back half of the Anger camera can be rejected as background-induced events. Furthermore, the rejection criteria can be

optimized for the energy of interest since the depth determination is performed *ex post facto* in the data analysis.

Figure 4.13 gives an indication of the depth determination capability of the camera plate. The detector was flooded from front and back by gamma-rays of 122 keV. Figure 4.13 shows the individual histograms of the depth parameter  $\lambda$  for events incident from the front and back. A depth selection which accepts 90% of the 122 keV gamma-rays incident from the front will reject 87% of the gamma-rays incident from the back. Alternatively, a depth selection which accepts 90% of the events incident on the front of the NaI(Tl) detector will reduce a uniform internal background by a factor of two.

The background rejection provided by the depth selection is capable of improving the flux sensitivity of the detector. There are however serious difficulties in realizing this improvement. The depth resolution varies dramatically in the plane of the detector. Above the center of a PMT the resolution is a few mm at 122 keV, while closer to the crystal face, there is less depth information. Near the intersection between three PMTs there is little or no depth information. Thus the effectiveness of the depth discrimination varies from point to point in the detector, complicating the analysis of the resulting data. In hindsight, the depth resolution could be improved and made more uniform by employing a larger number of smaller PMTs.

**Figure 4.13.** Depth discrimination. The figure shows histograms of the depth parameter  $\lambda$  for front and back floods of the detector at 122 keV. No restriction has been placed on the  $(x,y)$  positions of the events. Selection on the depth parameter allows 87% rejection of 122 keV events from the back on the detector, with 90% acceptance of events from the front. This discrimination has, however, a complicated dependence on the  $(x,y)$  position.





#### 4.2.3 The Calibration Systems

Careful measurement of changes in the gains of the primary detector PMTs and the gains and offsets of their associated electronics are needed in order for accurate reconstruction of the energy and position of events in the detector. For this purpose three calibration systems are employed. The first measures the offset and gains of each electronics channel, the second measures the response of the PMTs to light sources at fixed locations within the primary detector assembly, and the third measures the detector response to gamma-rays of a known energy.

On command, the processing of an event can be initiated without an energy deposit within the primary detector or shields. This allows the measurement of the baseline level of each channel. Offsets from the nominal null point can be caused by temperature-induced drifts in the electronics, changes in the dark current of the PMTs, or changes in the afterglow level of the NaI crystal. Experimentally we found that 10% of the light emitted in the NaI after an event is associated with a long ( $\sim 0.25$ s) time constant. Thus the crystal glows with a brightness proportional to the event rate. This glow causes an offset in the converted signals and also contributes to statistical noise.

Attached to the front of each pre-amplifier is a pulser that delivers a known charge with a current signal having the  $2.5\mu\text{s}$  time constant of the NaI crystal. Firing the pulsers and at the same time commanding the processing of an event produces data that, in conjunction with the offset data, may be used to measure the gain of each electronics channel.

A system of twelve green light-emitting diodes was installed in the detector assembly to allow the measurement of any relative changes in the gains of

the PMTs. A LED is located near the optical window in each junction between three PMTs. One diode is pulsed every  $\sim 15$ ms and the responses of the PMTs is recorded in the telemetry stream for later analysis. Insofar as the optical properties of the detector assembly and the ratio of the blue to the green response of the photocathodes do not change in time, the LED system allows a measurement of the ratios of PMT gains. Because we do not wish to rely on the stability of the LED light outputs the absolute gain calibration does not rely on them.

Absolute energy calibration of the primary detector is accomplished through the use of a radioactive source located above the mask. Located on a boom 1 m above the mask and 0.75 m from the axis of the telescope is 1  $\mu$ Ci of  $^{241}\text{Am}$  embedded in a plastic scintillator. Thirty-six percent of the alpha decays of the  $^{241}\text{Am}$  are accompanied by the emission of a 60 keV photon. The coincident photon is the result of a 68 ns half-life excited state of the  $^{237}\text{Np}$  created by the alpha decay. A photo-tube connected to the  $^{241}\text{Am}$  doped scintillator detects the alpha particle emission, allowing the 60 keV photons to be tagged. This source of tagged photons with known energy allows an absolute calibration of the primary detector's energy scale.

The known source location of the tagged photons provides a calibration of the imaging system. Since the calibration source has been placed above the mask, the tagged photons will be spatially modulated with the mask pattern. This modulation may be checked for consistency with the assumed orientation of the mask, allowing detection of any problems with mask position or distortion in the calculation of event positions within the detector. Detailed analysis of the calibration source events should also allow a check on the

relative PMT gains determined using the LED calibration system.

#### **4.2.4 Possible Design Improvements**

After a detailed examination of the response of the detector we realized that there were several areas where the design could have been improved. Primary among these was the deployment of PMTs near the edge of the detector. Near the edge there are regions with no PMT coverage. This results in regions with poor position resolution that must be excluded from use for imaging. It would have been more desirable to include these regions by using PMTs that extend over the edge of the NaI crystal. In the juncture between three PMTs in a region of poor sensitivity to depth. This sensitivity could have been improved by using an array of smaller hexagonal PMTs. Of course both of these changes involve additional electronics and increase the telemetry requirements. This latter difficulty could be avoided if on-board position calculation were implemented, but this is probably unwise until more experience is gained with the detector. The lack of depth sensitivity near the front face of the detector could most likely be eliminated by changing the nature of the reflector used on the surface of the housing. In particular, a "corner cube" type reflector might give a better response.

#### **4.2.5 The Shield System**

The primary detector is surrounded by a cylinder of twelve plastic scintillator modules as shown in Figure 4.1. The bottom of the cylinder is closed off with a NaI camera plate assembly which is identical to the primary detector assembly. This NaI shield serves as a backup to the primary detector,

and if called for the two detector assemblies can be interchanged.

One function of the shielding system is to lower the detector background due to external photons. This is accomplished at the expense of additional weight and instrumentational complexity, and if the only considerations are flux sensitivity and total weight, then it is possible that an unshielded detector would have been a better choice. This can be seen by an examination of the improvement in source flux sensitivity gained by the addition of an active shield. In a background-dominated measurement, the flux sensitivity is proportional to the square root of the ratio of the background rate per unit detector area to the detector area. A factor of two improvement in sensitivity may be gained by either decreasing the background by a factor of four, or increasing the detector area by a factor of four. Thus, if more than three times the mass of the detector in active shielding material is needed to decrease the background by a factor of four, this material is better utilized in the form of additional detectors.

There are additional considerations that make an active shield advantageous. The reduction of the background improves the signal-to-noise ratio of a source flux measurement. This reduces the size of any systematic errors. In addition, the shield serves to detect photons that escape from the detector after partial deposition of their energy. For the GRIP instrument a primary consideration was the telemetry requirement. Adding detector area to improve the flux sensitivity increases the event rate and so the required telemetry throughput, while reduction of the background with active shielding to improve the flux sensitivity reduces the event rate and so the telemetry requirement.

The cylinder of plastic scintillator is 96 cm in internal radius and  $\sim 16$  cm thick, and extends from 46 cm above to 45 cm below the top of the primary detector. The upper 39 cm of each module is tapered from the full thickness to 7 cm at the top. The modules were constructed by Bicron Corp. of type BC-408 scintillation plastic. Each module is wrapped in Millipore, an efficient diffuse reflector, and viewed from below by a single 5 inch Hamamatsu R1416 PMT.

For low-energy background photons the dominant mode of energy deposit differs between the NaI shield and the plastic shield. Below 250 keV a photon is more likely to interact in NaI via the photo-electric effect than to Compton scatter. However, in the lower Z plastic (mainly carbon) Compton scattering is the primary means of energy loss for photons with energy above 20 keV. Photons with energy less than  $\sim 100$  keV must Compton scatter in the plastic several times in order to produce a detectable signal. For photons entering the plastic from outside the cylinder this will in fact happen; however, it is possible for low-energy photons to scatter off the interior wall of the plastic shield without depositing a detectable amount of energy. To reduce this scattered flux we have lined the interior wall of the shield with a thin layer of tin. The tin layer is 1.6 mm thick and extends down 46 cm from the top of the plastic shield.

The analog sum of the NaI shield PMTs is amplified and fed to a discriminator with a commandable threshold. The event logic uses the output of this discriminator to inhibit processing of events when there has been a detected deposit of energy in the NaI shield. Event processing is inhibited while the PMT sum signal exceeds the threshold level and  $.125 \mu\text{s}$  afterwards.

Typically the total dead-time induced by a gamma-ray event in the NaI shield is less than 2  $\mu$ s.

The electronics associated with the plastic shield modules is slightly more complicated. The light emitted by a plastic shield module after the passage of an ionizing particle saturates the PMT. Full recovery from saturation takes 20 to 50  $\mu$ s depending on the energy deposit. The signal from the tube consists of a short ( $\sim 300$ ns) pulse followed by a slowly dropping tail with occasional small bursts. To allow full recovery from saturation without imposing excessive dead-time for gamma-ray events in the plastic shield, each module is instrumented with three discriminators, from which an inhibit signal is derived. The discriminator with the highest level is used to detect the large energy deposit events caused by charged particles. The other two discriminators detect the small energy deposits caused by gamma-ray interactions. The inhibit signal from a shield module is normally the output of the discriminator with the lowest threshold. However after the passage of a charged particle the inhibit signal is switched to the middle level discriminator. This allows a faster recovery from the saturation, at the expense of a less sensitive threshold. When the lowest level discriminator remains off continuously for 1.25  $\mu$ s the inhibit signal is switched back to the lowest level discriminator. If desired, this bi-leveled discrimination can be disabled. For each processed event the number of shield modules at the middle discrimination level is recorded, and the rate system measures the time spent in the less sensitive mode of operation.

The light collection characteristics of the plastic modules were tested using a  $^{137}\text{Cs}$  source. The light collection efficiency was found to vary by 8%

over the length of a module. The average response was 19 photo electrons collected per 30 keV of energy deposit, representing a light collection efficiency of approximately 25%.

Monte Carlo simulations of the shielded detector were performed to evaluate the performance of the shields. These showed a reduction in the background over that expected for an unshielded detector that varied from a factor of 7.4 at 100 keV to a factor of 1.7 at 2 MeV. Due to the lack of data taken in the same background environment with the shields present and removed, there is little with which to confirm these calculations. The data that exists consists of integral shield and detector rates, and the energy loss spectra in the primary detector with and without shield anti-coincidence.

The integral shield rates may be used to estimate the local background above the shield threshold. Data from the flight bay in Palestine, Texas, showed a shield counting rate of 30,000 counts/sec in the plastic shield with the threshold nominally set at 30 keV. Prediction of the counting rate in this same background for the unshielded primary detector is complicated by the non-isotropic nature of the background and the differing efficiencies between the plastic and NaI at 30 keV. We will however ignore these factors and scale the background rate by the ratio of the geometry factor of the unshielded primary detector to that of the plastic shield at low energies. This results in a predicted rate of 3700 counts/sec above 30 keV for the unshielded primary detector. The measured shielded rate is 1100 counts/sec with no shield anti-coincidence, and 600 counts/sec with shield anti-coincidence. This data therefore indicates that the total count rate has been passively reduced by a factor of  $\sim 3$ , with another factor of  $\sim 2$  reduction accomplished actively.

The integral rates depend mainly on the low-energy background, and reveal little about the background reduction at higher energy. The change in the background rate at any energy could be due to the passive effect of the shields, or the active effect of the shield's anti-coincidence veto. Due to the difficulty and risk associated with removing the detector from within the shield, we have no data on its passive effect. In any case we expect the passive effect of the shield to be markedly different in flight than on the ground, due to the creation of gamma-rays in the shield through interaction with the primary particle flux in the upper atmosphere, which caused a net increase in the background at higher energies.

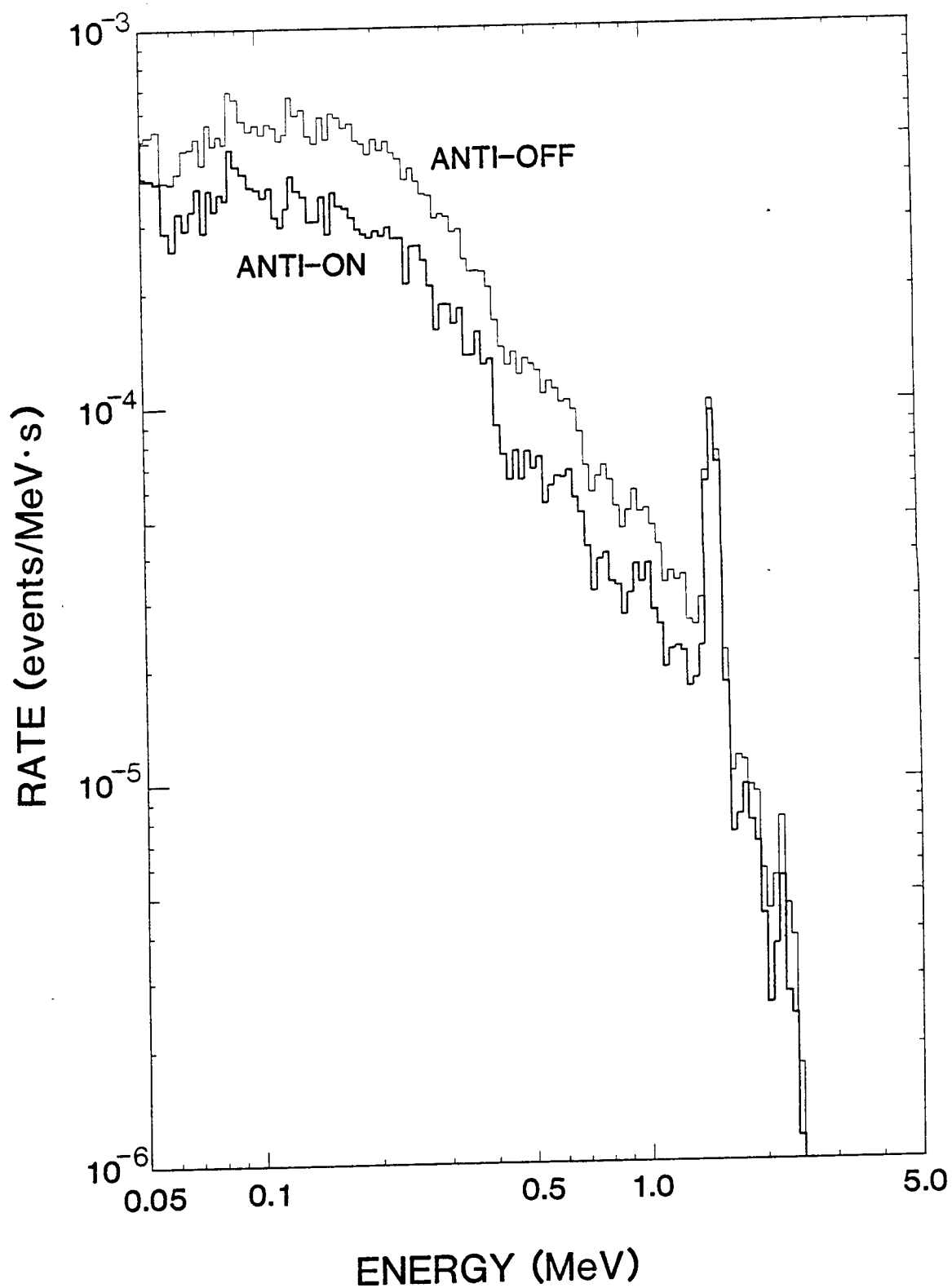
Figure 4.14 shows the effect of the anti-coincidence veto on the background in the flight bay at Palestine, Texas. The 'anti-off' spectrum is for all events within 14cm of the detectors center irrespective of the status of the shields. The 'anti-on' spectrum has those events with a coincident interaction in a shield removed. The prominent line feature at 1.4 MeV is due to naturally occurring  $^{40}\text{K}$ . The background at the line energy is not reduced by the shield anti-coincidence because most photons at this energy have not been scattered. However the continuum below this line is reduced by almost a factor of two by the shield anti-coincidence veto, and the amount of reduction is fairly independent of energy.

### 4.3 Support Sub-Systems

The basic imaging system described in section 4.2 is supported by numerous sub-systems for the processing of interactions and the control of environment and orientation. This section describes these sub-systems and



**Figure 4.14.** Anti-on and Anti-off spectrum. The figure shows energy loss spectra for the primary detector in the flight bay at NSBF in Palestine, Texas, in the fall of 1985. Events are restricted to the inner 14 cm of the primary detector. The Anti-off spectrum shows all events irrespective of the shield status, while the Anti-on spectrum includes only events with no coincident energy deposit detected in the shield.

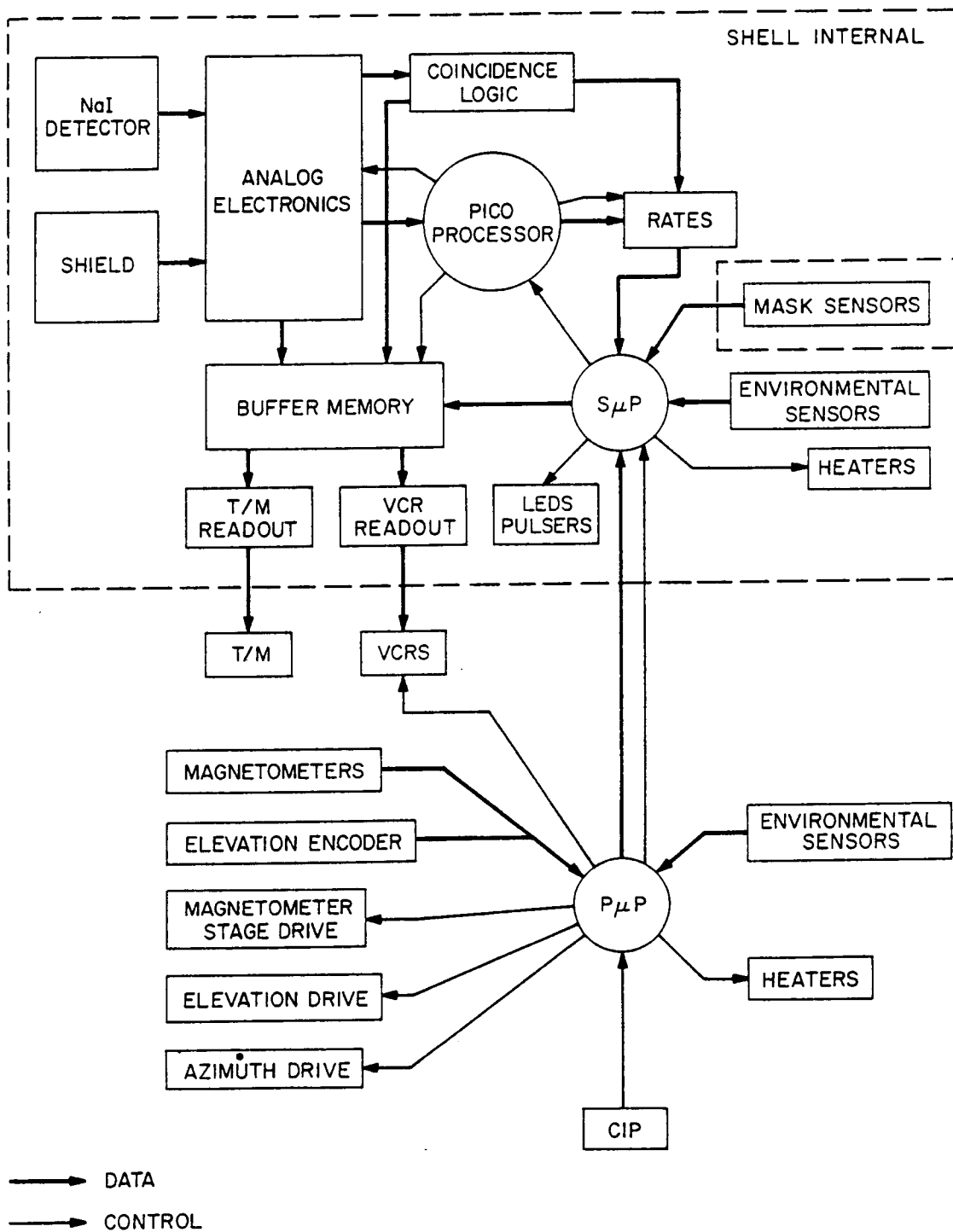


their organization.

Figure 4.15 shows a diagram of the overall organization of the GRIP telescope. For thermal and mechanical reasons the support systems are separated into two groups, one inside the shell or pressured enclosure that houses the primary detector and its active shield, and another outside this shell on the pointing platform. Control of the instrument is organized around two Rockwell micro-processors and a device we call a "pico-processor." One micro-processor is located outside the shell and is primarily responsible for the control of the instrument's pointing direction. The other micro-processor is located inside the shell and is primarily responsible for the control of the flow of data from the primary detector and active shields. These are called respectively the Pointing micro-Processor ( $P\mu P$ ) and the Shell micro-Processor ( $S\mu P$ ). The pico-processor resides inside the shell and is responsible for the processing of individual events.

The basis flow of data can be followed in Figure 4.15. The outputs of the detector and shield are monitored by analog electronics that produce discriminator signals used by the pico-processor to control the processing of events. When event processing has been triggered, the pico-processor checks whether coincidence requirements are met, and if so commands the conversion of the detector PMT pulse heights to digital values which are stored in a buffer memory. Blocks of events are combined in the buffer memory with house-keeping information supplied by the  $S\mu P$  and through the  $S\mu P$  by the  $P\mu P$ , and this data is transferred periodically to both a readout system for telemetry to the ground, and to a readout system for recording on video-tape. The coincidence requirements used by the pico-processor and the rates being

**Figure 4.15.** Organization of the GRIP instrument. The figure shows the sub-systems of the GRIP instrument and their interconnection. Double arrows indicate the flow on data, while single arrows indicate the flow of control signals. Also shown is the division of systems between the shell and external to the shell. The organizational structure is discussed in the text.



measured are supplied by the S $\mu$ P and can be commanded from the ground.

The P $\mu$ P monitors magnetometers to determine GRIP's orientation, and drives an elevation stepping motor and an azimuthal torque motor to correct this orientation according to a pointing plan. This plan consists of a 48-hour schedule of observations which may be altered by ground command. The P $\mu$ P also controls the Video Cassette Recoder (VCR) system, and mediates command communication with the ground through the Command Interface Package (CIP).

#### 4.3.1 Event Logic

Activity in the primary detector and shields is recorded in two forms. The first form consists of events in the system that meet commandable coincidence requirements. In this case the response of the nineteen PMTs in the primary detector are recorded along with coincidence and timing information. The second form consists of rates for various kinds of events.

The data is grouped into structures that are called frames, with one frame being recorded every 11.71 ms. Each frame consists of 64 blocks of 32 bytes. The first block of the frame contains housekeeping information such as a frame number, rate information, timing information, and data from various temperature and pressure sensors. Data from up to 63 individual events that meet the coincidence requirements are recorded one per block in the remainder of the frame. If more than 63 events occur during the frame interval the excess events are lost, although the total number of events is recorded in the housekeeping. This corresponds to a maximum recording rate of 5378 events per second.

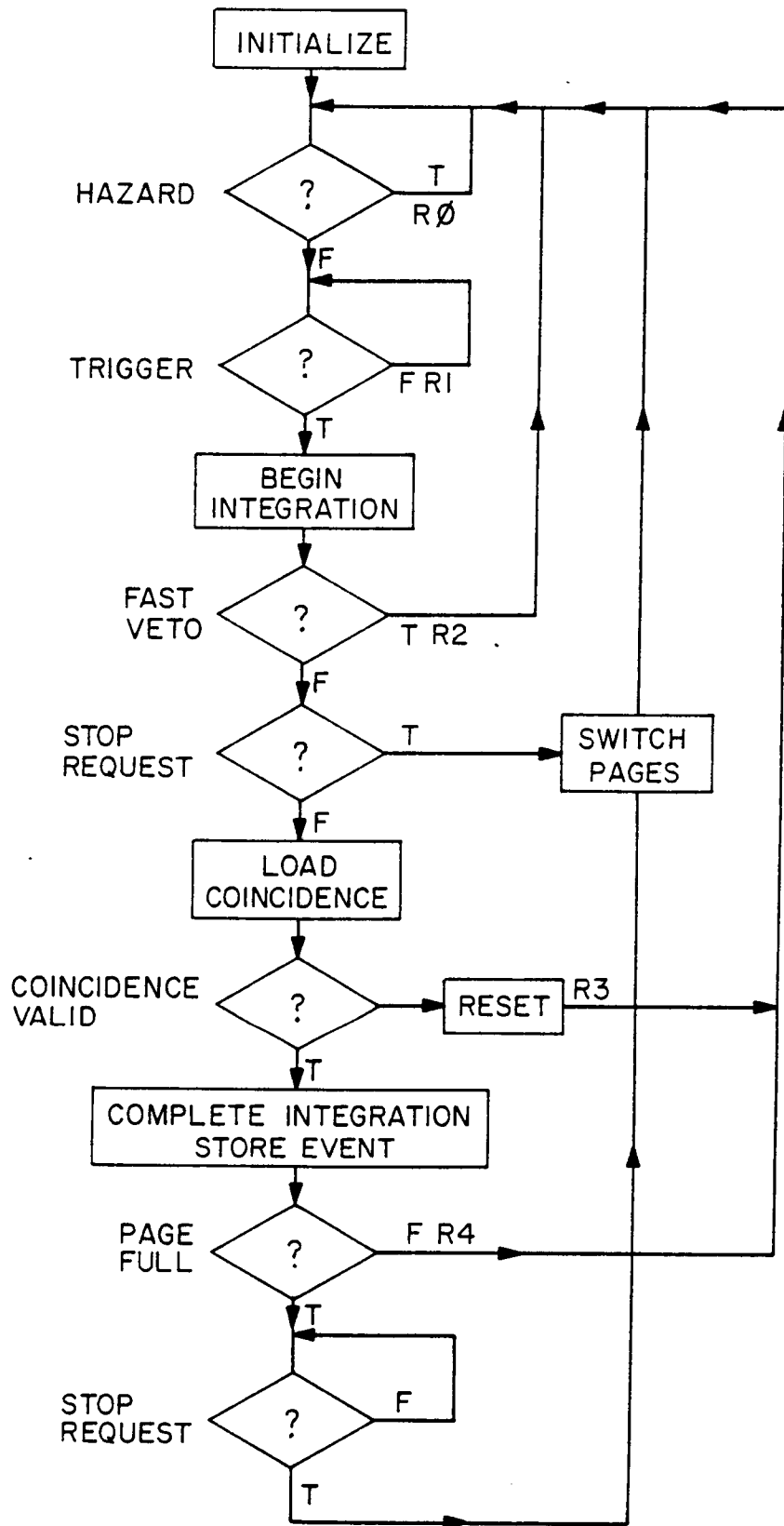
The processing of a frame of data takes three frame cycles. During the first cycle event data is loaded into a buffer memory. In the second cycle housekeeping information is loaded by the S $\mu$ P into the first block of the frame buffer. In the final cycle the frame is transferred to the telemetry and Video read-out systems for transmission to the ground and on-board recording. Three buffer memories are employed in a cyclic fashion to allow continuous transmission and recording of data.

The analysis and recording of events with valid coincidence is controlled by a device we call a "pico-processor." This is a finite state machine based on latched output Programmable Read Only Memories (PROMs). Five chips are employed to construct a single 9 bit address by 40 bit output PROM. At the beginning of the 125 ns clock cycle, the address lines are read, and then the 40 bit contents of that memory location are latched into the PROM's output buffer. Eight of the output bits are fed back to the address bus and determine all but the lowest order bit of the PROM address for the next clock cycle. Four output bits are used to address a multiplexer, which chooses one of 16 possible inputs as the lowest order bit of the PROM address for the next clock cycle. The remaining 28 output bits are discrete command lines that control various stages of the event analysis or increment rate counters. The contents of the PROM controls the timing and the flow of the event processing.

A portion of the flow chart for the program of the pico-processor is shown in Figure 4.16. At the beginning of a frame the pico-processor is in the initialization state, at the top of the flow chart. The flow follows in a straight line downwards with seven branches, all but one of which return to

**Figure 4.16.** The event logic flow chart. The flow chart summarizes the processing of events, which is managed by the pico-processor. Diamonds with question marks indicate branch points, with the branch taken being determined by the signal indicated to the left of the diamond. The rate counter signals associated with some branches are indicated by the labels R0 to R4. The event processing is described in detail in the text.





just after this initial state. For complete knowledge of the dead-time of the system, we need to record the time spent on any portion of the flow chart. Thus with five of the branch points is associated a rate signal. The remaining branch points involve the buffer memory switching which occurs at a fixed rate.

Immediately after initialization or the processing of an event the hazard signal is checked. This signal indicates that the system is not in a quiescent state, not having recovered from a recent shield or primary detector event. While this remains true the hazard signal is checked each cycle and rate 0 incremented.

After the hazard is cleared, the trigger signal is checked. The trigger signal indicates the detection of an interaction in the primary detector, the NaI shield or the plastic shield. The pico-processor remains in a loop checking the trigger signal and incrementing the live-time counter rate 1, as long as there is no trigger present. The system is live only when the pico-processor remains in this loop, so that rate 1 provides an accurate measure of the system live-time. This rate is counted and recorded every frame. All other rates are sampled in a commutation cycle by an additional counter. Immediately after receiving a trigger signal the pico-processor begins the integration of the pre-amplified signals from the 19 primary detector PMTs. This integration continues for  $1.5\mu\text{s}$  or until processing of the event is aborted.

After receiving a trigger the pico-processor checks the fast-veto signal. The fast-veto indicates that the event occurred exclusively in the shields with no energy deposit detected in the primary detector. If this is the case rate 2 is incremented and the pico-processor returns to the hazard loop, where it

remains until the system recovers from the event. Events that are restricted to the plastic shield cause a total dead-time of  $\sim 375$  ns while events in the NaI shield cause a dead-time of  $1.25 \mu\text{s}$ . Events which cause an energy deposit in the primary detector, but that do not meet the coincidence requirements, cause a larger dead-time. The fast-veto was introduced to prevent shield-only events from sharing this larger dead-time.

The next step in the flow chart involves the switching of buffer memory pages. At the end of a frame interval the micro-processor sends the stop request signal and forces the hazard and fast-veto signals low and the trigger signal high. Thus the pico-processor falls through the hazard and trigger loops and bypasses the fast-veto branch to arrive at the stop request branch. The presence of the stop request signal causes the pico-processor to execute the page switch sequence which takes 14 cycles to complete. After completion of the page switch sequence the pico-processor returns to the hazard loop.

For processing an event with energy deposit in the detector the stop request is normally off, and the pico-processor continues by next loading the coincidence signals into a buffer. This normally takes one cycle. These signals indicate the status of the primary detector, NaI shield and plastic shields. The signals are then checked for agreement with the coincidence requirements. The coincidence requirements are commandable, and can be altered by the micro-processor at the beginning of any frame interval. Normally an energy deposit is required in the primary detector, and no detectable energy deposit in any of the shields. If the coincidence is invalid, the pico-processor executes a reset routine which takes 6 cycles, increments rate 3 and returns to the hazard loop.

If the coincidence is determined to be valid, integration of the primary detector PMT signals is completed, and the signals are held and converted to a digital value. Each of the 19 primary detector PMTs has associated with it a 4096 channel A/D converter, allowing the conversions to occur in parallel. The converted signals are then loaded into the frame buffer along with the coincidence signals and a timing counter with 4  $\mu$ s resolution.

After storing the data from the processed event the pico-processor checks to see if the buffer memory is full. This occurs after the 63rd event in the frame interval. If the memory is not full, the pico-processor returns to the hazard loop. The total time from the initial trigger to return to the hazard loop is 12. $\mu$ s.

If the memory is full, the pico-processor enters a loop checking for the stop request signal. Thus no events are processed until the next frame interval. After a stop request the pico-processor executes the page switching routine and returns to the hazard loop.

#### **4.3.2 The Rate System**

Two counters are used to measure the rates of occurrence of a wide variety of events. The first counter is dedicated to measuring the live-time, while the second is used for all other single coincidence rate measurements. The S $\mu$ P selects the rate to be measured for each frame interval. At the beginning of the frame the counter is zeroed and connected to the selected signal. At the end of the frame the counter is loaded into the frames house-keeping block. The rates are at present commutated in a cycle of 256. Each signal may be measured in two modes, either measuring the number of times

the signal turned on, or the total time that it was on. The selectable signals include each of the three discriminators from each of the twelve plastic shield modules, the discriminator on the NaI shield, the ZLD, LLD and ULD of the primary detector and the discriminator on the calibration source. Thus the behavior of any active component of the system may be fully monitored.

#### **4.3.3 The Telemetry and Recording Systems**

The GRIP instrument generates data at a peak rate of 1.4 Mbits per second. The instrument has the on-board capability to record data at this rate for up to 40 hours. In addition a limited portion of the data is transmitted to the ground.

On-board recording is essential because of the limitations of available telemetry systems. The telemetry link provided by the National Scientific Balloon Facility (NSBF) is limited in bandwidth and range. The bandwidth limitation allows the transmission of only the housekeeping data and the first 15 events of each frame. This signal is reliable only within a range of 300 miles from the balloon base. It provides data for real time analysis and trouble shooting, and acts as a backup to the on-board recording system.

A Video Cassette Recorder (VCR) system was developed for on-board recording of the GRIP data (Althouse and Cook 1985). The system combines commercially-available products to provide 40 hours of recording capacity at 1.4 Mbits per second with a power consumption of only 10 watts. The key components of the recording system are eight portable VCRs and a digital audio processor.

The digital audio processor (PCM-701ES) is a consumer product intended to allow the use of a home VCR as a digital audio recorder. In normal use incoming left and right audio signals are digitized with 16 bit resolution at a rate of 44kHz. The digital data is then input on a serial line to the "recording data processor" at a rate of 1.4 Mbits per second. The recording data processor adds error correcting codes and generates a video signal in standard NTSC format suitable for recording on a home VCR. On playback the "playback data processor" decodes the signal, corrects errors, and produces digital data used by 16 bit DACs to reconstruct the audio signals.

The GRIP instrument employs the digital board from the PCM-701ES. This has been modified to take as input a serial stream generated from the frame buffer memories. The NTSC video signal from the recording data processor is fed to one of eight Sony SL2000 portable VCRs. The VCRs which are controlled by the P $\mu$ P, are operated sequentially. Each has a recording capacity of five hours. The VCRs are housed in a pressurized thermally controlled vessel located on the pointing platform.

On the ground a modified digital board from the PCM-701ES is used to reconstruct the digital data from the recorded video signal. Testing of the recording system showed 30 frames per million must be rejected due to uncorrectable errors, a negligible loss of data for our purposes.

#### **4.3.4 The Pointing System**

The GRIP telescope is mounted on a two-axis pointing platform. This is shown in Figure 4.2. The elevation axis bearings are attached directly to the pressure vessel, near the center of mass of the telescope. The elevation of the

telescope is controlled by a ball-screw drive. The ball-screw, driven by a stepping motor, is mounted on the pointing platform, below and to the side of the telescope. The rotation of the ball-screw drives a linking rod which is attached near the bottom of the telescope. This drive system allows the adjustment of the telescope's elevation in steps of less than 3 arc minutes. The elevation is independently measured by a 6000 lines-per-revolution shaft encoder.

As is shown in Figure 4.2, the pointing platform is suspended by straps from a spreader bar, which is attached through a bearing to the helium balloon's suspension ladder. A torque motor is placed at this bearing, allowing control of the telescope's azimuthal orientation. The torque motor is capable of delivering 3 ft-lb maximum torque.

The azimuthal orientation of the pointing platform is measured by two pairs of magnetometers. Figure 4.2 shows their location on the pointing platform. The first pair measures the component of the magnetic field in the plane of the pointing platform. Each of the two magnetometers in this pair have a linear range that encompasses the full variation of Earth's magnetic field. The second pair consists of two Schonstedt MND-5C-25 magnetometers. These more sensitive magnetometers are within their linear range only if their axis is within  $5^\circ$  of being perpendicular to the magnetic field. These magnetometers are mounted on a 9000 steps-per-revolution rotation stage. One magnetometer has its axis in the plane of the platform, while the other has its axis perpendicular to the first and  $29^\circ$  from the plane of the platform, complementing the magnetic dip angle expected in flight. The pair is oriented during flight so that they are nulled when the platform is in the desired

orientation.

The P $\mu$ P is used to control the ball-screw stepping motor and the azimuth torque motor. This  $\mu$ -processor contains a schedule of the planned observation program which it uses to determine the correct orientation of the telescope relative to the magnetic field and the vertical. The elevation is controlled with no feedback from the elevation shaft encoder. The azimuth control normally uses feedback from the fixed orientation magnetometers, although it is possible to control with feedback from the Schonstedt magnetometers. The P $\mu$ P sends the outputs from the magnetometers and the shaft encoder, along with various temperatures and other measurements, to the S $\mu$ P for inclusion in the housekeeping data.

The pointing system is capable of operating in a robot mode without any external input for a complete 48-hour flight. Alternately the system may be commanded from the ground in a manual mode. Any portion of the observation schedule may be altered in flight in order to take advantage of sources of opportunity.

#### **4.3.5 Thermal Control Systems**

Considerable attention was paid to the thermal design of the GRIP instrument. During a balloon flight an instrument is exposed to wide range of thermal environments, and poor thermal control can lead to the failure of an otherwise well-designed instrument.

Much of the thermal control of the GRIP instrument is accomplished passively. At float altitude ( $\sim 120,000$  ft) there is little direct conductive or convective coupling of the instrument with the thin atmosphere. Instead the



instrument is primarily radiatively coupled to its environment. In this situation the heat that flows out through the surface of the instrument is the difference between the thermal emission of the surface and the combined inputs of direct and albedo solar radiation, and infrared radiation from the earth and atmosphere. Passive thermal control of the instrument can therefore be accomplished by proper selection of the ratio of surface area to heat generation, insulation and choice of surface materials and colors. By far the largest and most variable component of the heat input to the instrument surface is the solar radiation. We therefore decided that the surface of the instrument should be painted a bright white to minimize the solar input.

Preliminary estimates showed that the 300 watts dissipated by the GRIP instrument could not be radiated by a single pressure vessel of reasonable surface area. The electronics was therefore divided into three groups. The pressure vessel or shell contains the primary detector and shields with their support electronics. The electronics box is an unpressurized container on the pointing platform that contains all remaining electronics, except for the power converters which are directly coupled to the frame of the pointing platform. This grouping is quite natural, placing the thermally most sensitive components in the shell, the less sensitive in the electronics box, and coupling the rugged power converters to a large thermal radiator.

Computer models were used to determine how much insulation should be used on the shell and electronics box. The amount of insulation determines how closely the exterior surface follows the interior temperature. With little insulation the surface temperature is close to the interior temperature and the instrument surface radiates large amounts of heat. With too much insulation

the exterior surface temperature is near equilibrium (about  $-40^{\circ}\text{C}$  at night and  $0^{\circ}\text{C}$  at noon) and heat builds up in the package. We chose the insulation thicknesses so that thermal balance during the day was achieved at a reasonable temperature ( $20^{\circ}\text{C}$  for the shell and  $30^{\circ}\text{C}$  for the electronics box). Small (25 watt) heaters were included in each package to compensate for the additional heat loss expected at night.

Temperature control is critical for the NaI detectors, which may be fractured by thermal gradients. In addition the light response of the NaI and the gains of the PMTs are temperature sensitive. The NaI detectors were provided with heaters for emergency situations, and insulated from the plastic shields. The high voltage for the PMTs is generated externally to minimize heat sources near the NaI. Additional thermal buffering is provided by the large thermal mass of the plastic shields that surround the NaI detectors. This entire mass with enclosed NaI detectors was insulated from the remaining space within the shell. To distribute the heat generated by the shell electronics evenly over the pressure vessel surface a forced air convection system was employed.

Thermal control systems were also provided for various instruments located outside the main pressure vessel and electronics box. These included the pressure vessel housing the VCRs, the magnetometers and their rotation stage, the mask sensors, and the elevation encoder. Suitable insulation and regulated heaters were provided in each case.

To monitor the shell, electronics box, and other thermal control systems, the GRIP instrument includes a total of 32 temperature sensors. The responses of these sensors are measured along with other environmental data

in a commutation cycle and included in the housekeeping data. This information is therefore immediately available during a flight, and recorded for post-flight analysis.

The thermal environments experienced at the beginning of a balloon flight are also important to consider. Before launch the instrument is often required to sit on the launch pad in full operational condition. On a hot sunny day in Texas an hour of fully enclosed operation on the launch pad may seriously over-heat the instrument. Then during ascent the instrument will be subjected to severely cold temperatures ( $-70^{\circ}\text{C}$ ) near the tropopause where there is good conductive and convective coupling to the atmosphere. For GRIP short exposures to these environments are not a problem, mainly due to the long thermal time constants of the instrument.

## V. Epilogue

I have argued that future low-energy gamma-ray observations can make important contributions to astronomy, but that the instruments that perform these observations will need finer angular resolution and better control of systematics than possible with the current generation of instruments. One possible technique for achieving finer resolution and better control of systematics is the coded aperture imaging technique analyzed in chapter 3 and employed in the GRIP instrument, described in chapter 4. The maiden flight of the GRIP instrument provided a test of this imaging technique, and will be the subject of this chapter.

However before discussing this flight, I must digress to discuss my past and present involvement with the GRIP instrument. This began with participation in design studies of gamma-ray instruments, when I started research in the Space Radiation Laboratory at Caltech in 1980. These studies resulted in a tentative design of GRIP in 1982. During this time we conducted a laboratory test of coded aperture imaging, (Cook et al. 1984), which proved the veracity of the technique. Development of GRIP began soon thereafter, and the instrument was complete and operational in the fall of 1985. An attempt was made to launch GRIP on a high-altitude balloon from the National Scientific Balloon Facility in Palestine, Texas, during the jet-stream turnaround in the fall of 1985. I had hoped to obtain data for my thesis from this flight. However, no flight was made at that time, due to problems with

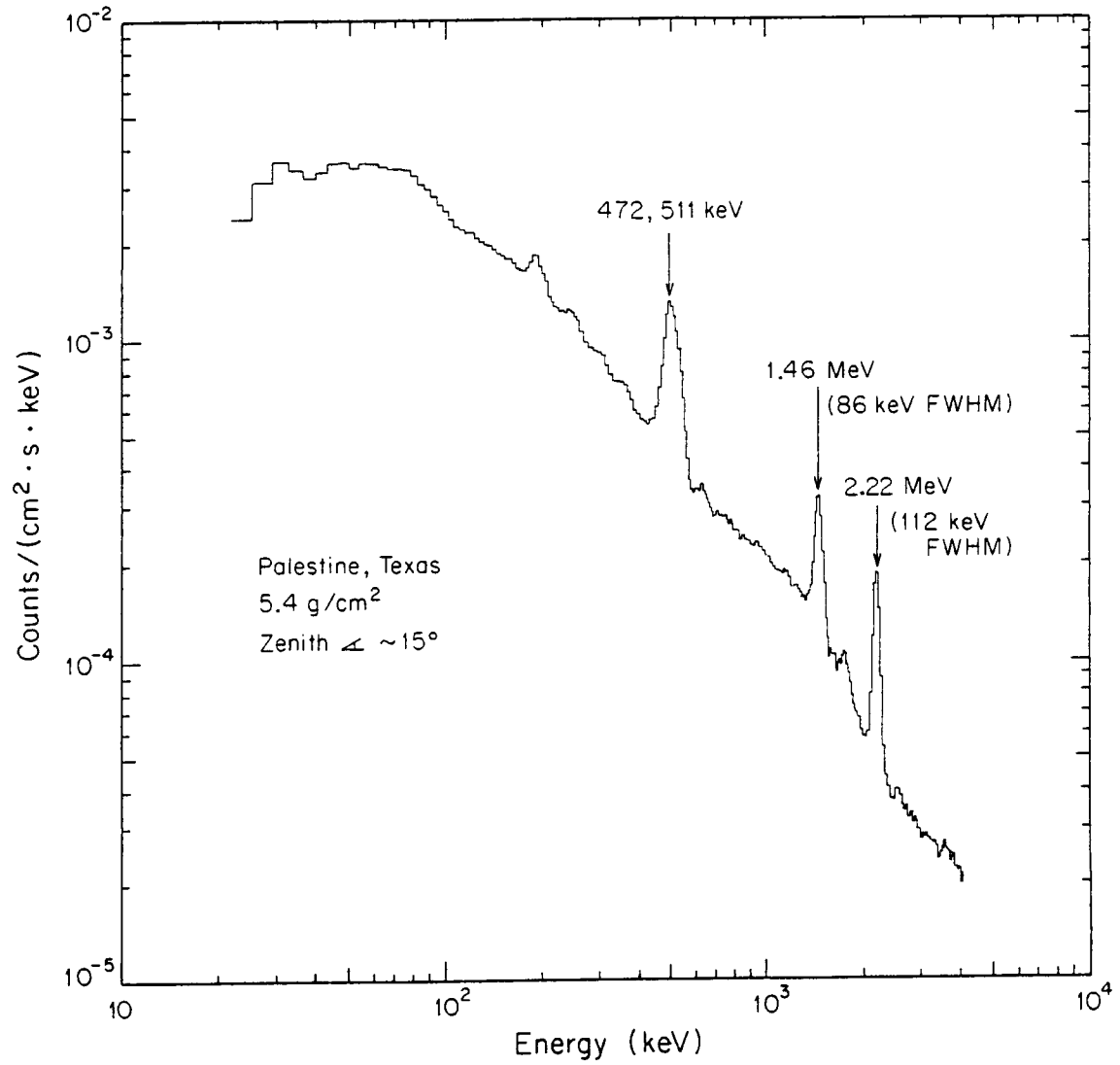
surface weather conditions and balloon materials. The balloon material problems were not resolved in time for the spring turn-around of 1986. GRIP was launched for its first flight on October 15, 1986, from Palestine, Texas. It flew successfully for 28 hours and landed in the south-east corner of Arkansas with minimal damage. By this time I had decided not to rely on flight data for my thesis, and was therefore not directly involved with this flight, or the subsequent data analysis.

During GRIP's maiden flight (Althouse et al. 1987), pointed observations were made in the directions of the quasar 3C273, the Cygnus region, the galactic center region, the COS-B source GC135+1, and the Crab Nebula. Constant monitoring of the housekeeping information showed that the instrument operated continuously and according to plan. Confidence in the functionality of the instrument was bolstered by the successful production of a real-time image of the black hole candidate Cygnus X-1. Although detailed analysis of the instrument's performance is still in progress, all indications show that the flight was an unqualified success.

Figure 5.1 shows the spectrum of the detector background during flight. The line feature at 2.2 MeV is due to neutron capture on protons in the plastic shield. The line at 1.46 MeV is due to  $^{40}\text{K}$ , probably in material near the instrument since this line is seen at nearly the same intensity on the ground. The line near 0.5 MeV is thought to be a blend of the 0.511 MeV positron annihilation line with a line at 0.472 MeV due to neutron scattering on Na in the detector. It is apparent that a step in the continuum appears below each of these lines, indicating the presence of Compton tails. From this background spectrum we may deduce the instrument's flux sensitivity. Proceeding

**Figure 5.1.** The background energy loss spectrum during the first flight of the GRIP instrument in October of 1986. The spectrum was taken at  $5.4 \text{ g/cm}^2$  of atmospheric depth with the telescope pointed near  $15^\circ$  from the zenith. The identified lines and their widths are discussed in the text.

### In-Flight Background Spectrum



from equation (3.29) and including factors for the detector efficiency  $\epsilon$ , the mask opacity  $\rho$ , the energy interval  $\Delta E$ , and atmospheric attenuation  $\alpha$ , the  $3\sigma$  continuum sensitivity is given by,

$$F_{3\sigma} = \frac{6}{\epsilon\alpha\rho} \sqrt{\frac{B}{A\tau\Delta E F_z}} \quad (5.1)$$

where  $B$  is the background spectrum given in figure 5.1, and  $A \approx 600 \text{ cm}^2$  is the usable area of the detector. Assuming a duration  $\tau$  of four hours, I estimate a sensitivity of  $\sim 2 \cdot 10^{-5} \text{ photons/cm}^2 \cdot \text{s} \cdot \text{keV}$  at 100 keV, and at 1 MeV a sensitivity of  $\sim 3 \cdot 10^{-6} \text{ photons/cm}^2 \cdot \text{s} \cdot \text{keV}$ , comparable with other balloon-borne instruments, but modest compared to the needs outlined in chapter 1. Clearly the lower background obtainable in space, along with the prolonged exposures possible, will be needed to detect a large number of extra-galactic sources.

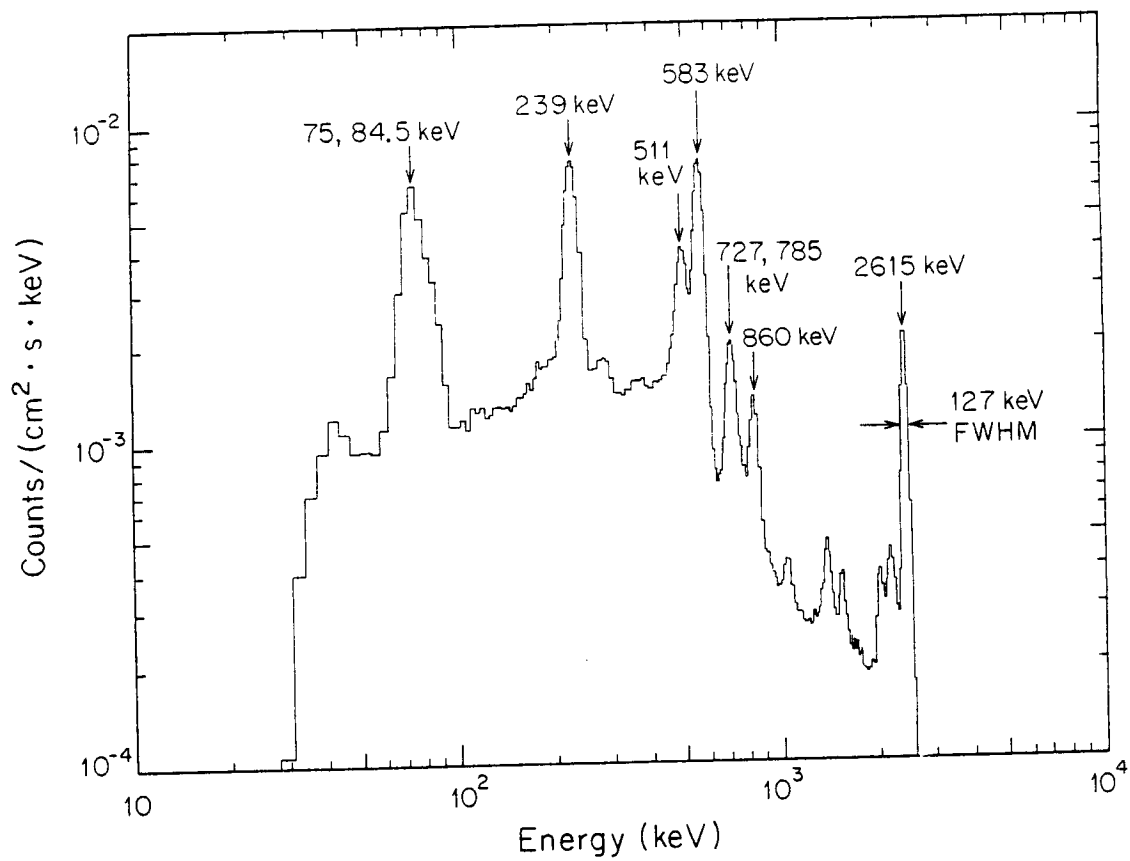
The energy resolution indicated in figure 5.1 for the 1.46 and 2.22 MeV lines is consistent with results obtained on the ground. Fig 5.2 shows the energy spectrum measured in the laboratory for using a  $^{228}\text{Th}$  source. The good energy resolution has been achieved with this large detector by detailed mapping of light collection efficiency. This resolution is maintained in flight by use of the LED and  $^{241}\text{Am}$  source calibrations which allow continuous monitoring of the PMT gains.

Analysis of the data from the first GRIP flight is still in progress, however preliminary images are available for the Crab Nebula and Cygnus regions. These are shown in figures 5.3 and 5.4 in the form of contour plots of the image  $C(\bar{z})$  over the  $20^\circ$  fully modulated field of view. Dashed lines

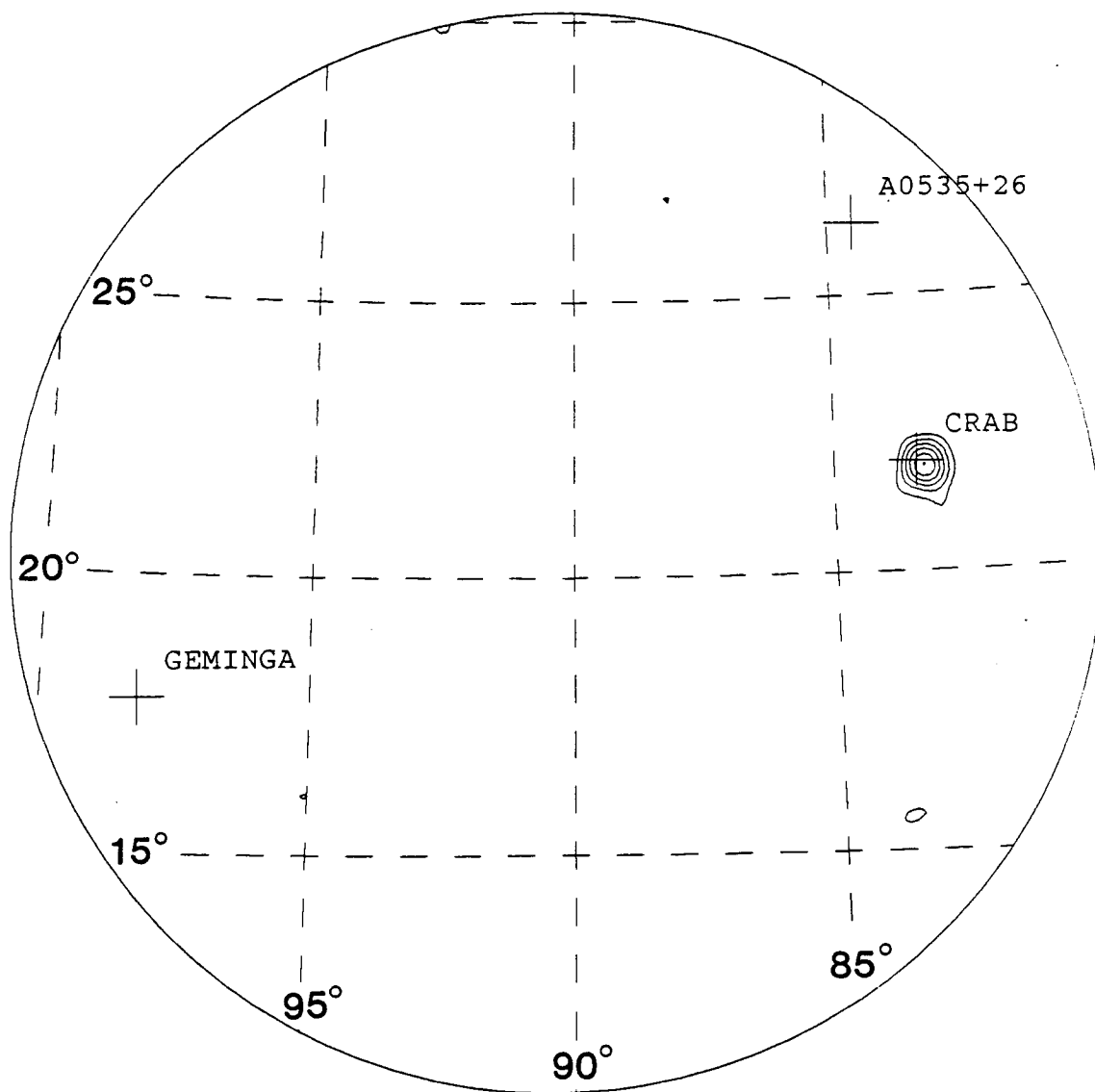


**Figure 5.2.**  $^{228}\text{Th}$  spectrum obtained with the primary NaI(Tl) detector demonstrating the energy resolution obtained on the ground. Energies are identified for the strongest lines.

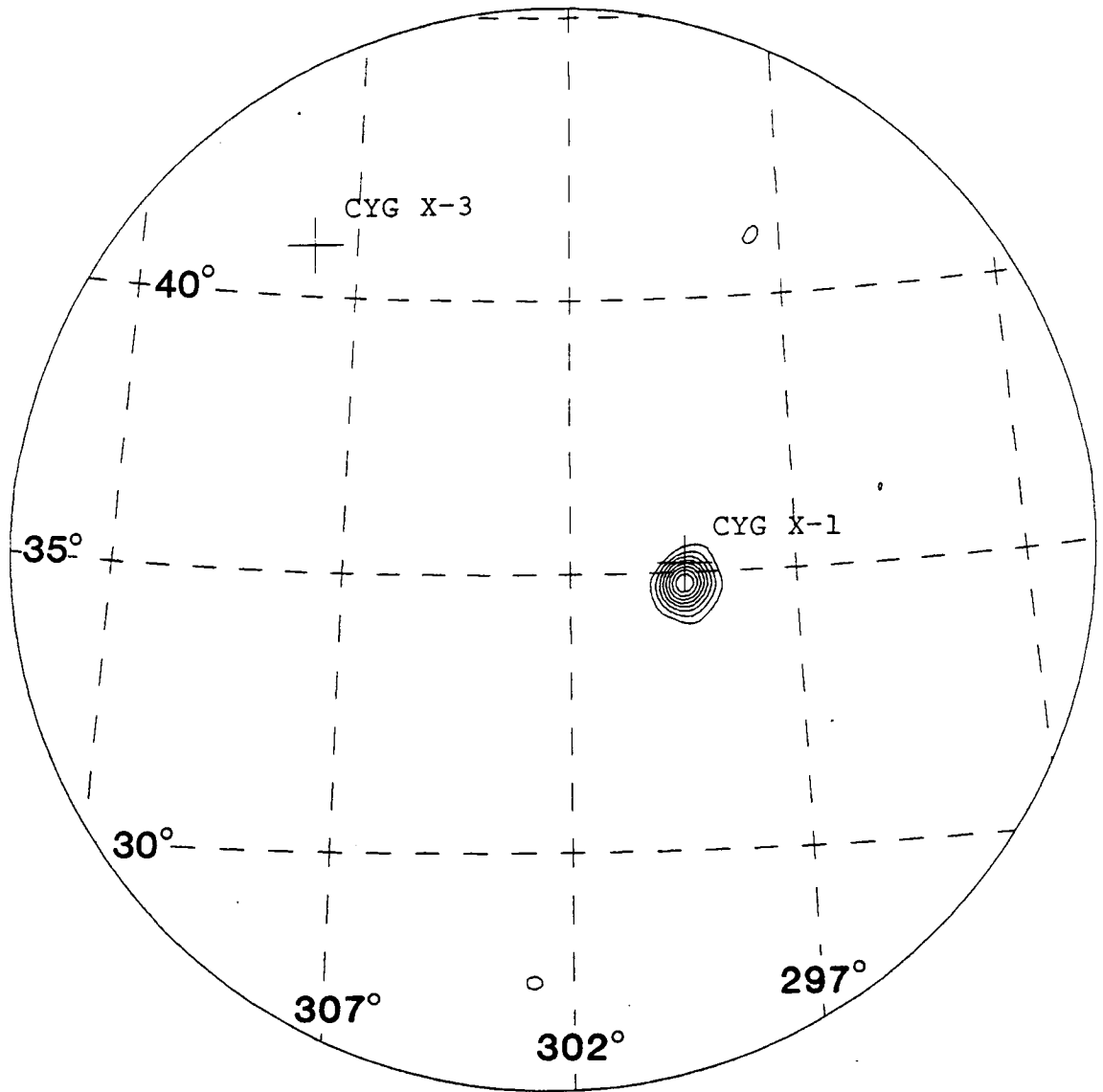
$^{228}\text{Th}$  Laboratory Spectrum



**Figure 5.3.** Image of Crab Nebula regions produced from GRIP observations. Contours indicate levels of statistical significance and have  $2\sigma$  spacing. Noise peaks at the lowest contour level have a significance of  $3\sigma$ . Crosses indicate the actual locations of candidate sources, with the dashed lines indicating right ascension and declination.



**Figure 5.4.** Image of Cygnus regions produced from GRIP observations. Contours indicate levels of statistical significance and have  $2\sigma$  spacing. Noise peaks at the lowest contour level have a significance of  $3\sigma$ . Crosses indicate the actual locations of candidate sources, with the dashed lines indicating right ascension and declination.



indicate right ascension and declination coordinates, while cross-bars indicate the locations of sources within the fields-of-view. The images are based on two hours of data in the 50-150 keV energy band. At present the flux scale is not fully calibrated. Therefore the image contours are in units of statistical fluctuation level  $\sigma$  of the image. These contours start at a  $3\sigma$  level and increase in  $2\sigma$  steps. The Crab is detected at the  $13\sigma$  level, while Cygnus X-1 is detected at the  $17\sigma$  level. At present the candidate sources Cygnus X-3, Geminga, and A0535+26 have not been detected. Work to detect or establish upper limits for these sources is in progress. The offsets of the image peak locations from the actual locations of the detected sources is consistent with the present estimate of the pointing uncertainty. As analysis of the data proceeds, these pointing errors should be significantly reduced.

It should be noted that the Crab image contains 4 noise peaks above the  $3\sigma$  level, while the Cygnus image contains 3 noise peaks above the  $3\sigma$  level. This is consistent with the expected number of  $\sim 5$  noise peaks above  $3\sigma$ , which can be deduced from figures 3.6 and 4.12.

Non-imaging gamma-ray astronomers may be horrified by this large number of  $3\sigma$  fluctuations. However this is merely a reflection of the large number of flux measurements that are actually being presented in each image. Herein lies the power of the coded-aperture imaging technique for the control of systematics. The degree of consistency of the non-source portions of the image with a uniform flux provides a measure of the level of systematic errors, and any systematic trends in the image noise provide clues to the sources of these errors.

## Appendix: Properties of the Image Function

In this appendix we derive certain useful properties of the image function that follow from the HURA property of the mask, the use of mask-anti-mask subtraction, and the use of continuous position sensitive detectors. We start with the definition of the image function given in equation (3.20):

$$f(\bar{z}, \bar{z}') = \frac{1}{A} \int_A \Delta M^{eff}(\bar{z} + \bar{z}) \Delta M^{eff}(\bar{z} + \bar{z}') d^2x \quad . \quad (\text{A.1})$$

Because the integration is over a full repetition of a periodic function, we can change variables to  $\bar{x}' = \bar{z} + \bar{z}'$  without translating the area of integration. By changing integration variables alternately to  $\bar{x}' = \bar{z} + \bar{z}'$  and to  $\bar{x} = \bar{z} + \bar{z}$  it can be seen that the result depends only on  $\bar{z} - \bar{z}'$ , and further that the result is independent of the sign of  $\bar{z} - \bar{z}'$ :

$$f(\bar{z}, \bar{z}') = f(\bar{z} - \bar{z}') = f(\bar{z}' - \bar{z}) \quad . \quad (\text{A.2})$$

The image function thus has inversion symmetry.

Other *approximate* symmetries of the image function are useful. In particular, for an HURA, the image function is azimuthally symmetric up to 4th order in  $\bar{z}$ . This can be most easily seen if  $f(\bar{z})$  is expanded as a general polynomial of the form:

$$\sum_{jk} \alpha_{jk} (\zeta)^j (\zeta^*)^k = \sum_{jk} \alpha_{jk} r^{j+k} e^{i\phi(j-k)} \quad (\text{A.3})$$



with  $\zeta = z_x + i z_y = r e^{i\phi}$  and  $\zeta^* = z_x - i z_y$ . The hexagonal symmetry of the image function (see below) forces all coefficients,  $\alpha_{jk}$ , to vanish unless  $j-k$  is a multiple of 6. Thus for  $|j+k|$  less than 6 we must have  $j=k$  and  $f(\bar{z})$  has azimuthal symmetry for order less than 6. This implies that all derivatives of  $f(\bar{z})$  at  $\bar{z}=0$  of a given order less than 6 are multiples of a single value. Note that applying the same reasoning to rectangular URAs, we find that deviations from azimuthal symmetry may enter in the 4th order terms.

We now derive a simple and particularly useful form for the image function  $f(\bar{z})$ . We first decompose the mask function into its hexagonal cells :

$$M_{\pm}(\bar{y}) = \sum_{j=0}^{v-1} m_j^{\pm} h(\bar{y} - \bar{c}_j) \quad (\text{A.4})$$

where  $m_j^{\pm}$  is zero or one,  $h(\bar{y})$  is a periodic function that is one inside the central cell and its repetitions and zero elsewhere, and  $\bar{c}_j$  is a cell center inside the unit pattern. The effective mask difference function can then be expressed as

$$\Delta M^{eff}(\bar{y}) = \sum_j n_j h^{eff}(\bar{y} - \bar{c}_j) \quad (\text{A.5})$$

where

$$h^{eff}(\bar{y}) = \int p(\bar{\Delta}) h(\bar{y} - \bar{\Delta}) d^2 \Delta \quad (\text{A.6})$$

and where  $n_j = m_j^+ - m_j^-$  is plus or minus one except for the central cell for which it is zero, and  $p(\bar{\Delta})$  is the detector point spread function. Substituting equation (A.5) into equation (A.1) we have

$$f(\bar{z}) = \sum_{jk} n_j n_k \frac{1}{A} \int_A h^{eff}(\bar{x} - \bar{c}_j) h^{eff}(\bar{x} + \bar{z} - \bar{c}_k) d^2 x \quad (A.7)$$

which upon changing integration variables to  $\bar{x}' = \bar{x} - \bar{c}_j$  results in

$$f(\bar{z}) = \sum_{jk} n_j n_k \frac{1}{A} \int_A h^{eff}(\bar{x}') h^{eff}(\bar{x}' + \bar{z} - [\bar{c}_k - \bar{c}_j]) d^2 x' \quad (A.8)$$

Because the effective hexagon function  $h^{eff}(\bar{y})$  has the periodicity of the mask, any difference  $\bar{c}_j - \bar{c}_k$  is equivalent to one of the  $v$  cell centers  $\bar{c}_l$  chosen from the central unit pattern. Thus we have

$$f(\bar{z}) = \sum_{l=0}^{v-1} b_l \frac{1}{A} \int_A h^{eff}(\bar{x}') h^{eff}(\bar{x}' + \bar{z} - \bar{c}_l) d^2 x' \quad (A.9)$$

where

$$b_l = \sum_{jk} n_j n_k \delta_{\bar{c}_k - \bar{c}_j, \bar{c}_l} \quad (A.10)$$

At this point we may use the uniform redundancy property of the mask pattern to compute the sums  $b_l$ . For a given non-zero difference  $\bar{c}_l$ , by the uniform redundancy property there are  $(v-3)/4$  opaque cells separated by  $\bar{c}_l$  from opaque cells in the unrotated mask. Thus in the sum for  $b_l$  there are  $(v-3)/4$  terms with  $n_j = n_k = -1$ . By the anti-symmetry of the mask we can conclude there are  $(v-3)/4$  non-central transparent cells separated by  $\bar{c}_l$  from non central transparent cells in the unrotated mask. Thus there are  $(v-3)/4$  terms with  $n_j = n_k = 1$ . For the remaining  $(v+3)/2$  terms we have  $n_j \neq n_k$ . Of these, two involve the central cell with  $n_j$  or  $n_k$  being zero. For the rest we have  $n_j n_k = -1$ . Thus we have

$$b_l = v \delta_{l,0} - 1 \quad . \quad (\text{A.11})$$

We have incorporated here the result for  $b_0$  corresponding to a zero difference. With this result, and noting that the sum of all  $v$  hexagon functions  $h^{eff}(\bar{y} - \bar{c}_l)$  results in a constant value of one, we have

$$f(\bar{z}) = \frac{v}{A} \int_A h^{eff}(\bar{x}) h^{eff}(\bar{x} + \bar{z}) d^2 x - \frac{1}{v} \quad . \quad (\text{A.12})$$

## References

- Althouse, W. E. and W. R. Cook, "Balloon-Borne Video Cassette Recorders for Digital Data Storage," *Proceedings of the 19th International Cosmic Ray Conference*, **3**, 395-397, 1985.
- Althouse, W. E., W. R. Cook, A. C. Cummings, M. H. Finger, D. M. Palmer, T. A. Prince, S. M. Schindler, C. H. Starr, and E. C. Stone, "First Flight of a New Balloon-Borne Gamma-Ray Imaging Telescope," submitted, 20 International Cosmic Ray Conference, 1987.
- Avni, Y., "On the Contribution of Active Galactic Nuclei to the Diffuse X-Ray Background," *Astronomy and Astrophysics*, **63**, L13-L16, 1978.
- Baity, W. A., R. E. Rothschild, R. E. Lingenfelter, W. A. Stein, P. L. Nolan, D. E. Gruber, F. K. Knight, J. L. Matteson, L. E. Peterson, F. A. Primini, A. M. Levine, W. H. G. Lewin, R. F. Mushotzky, and A. F. Tennant, "Centaurus A (NGC 5128) At 2 keV-2.3 MeV: HEAO-1 Observations and Implications," *Astrophysical Journal*, **244**, 429-435, 1981.
- Baity, W. A., R. F. Mushotzky, D. M. Worrall, R. E. Rothschild, A. F. Tennant, and F. A. Primini, "Observations of NGC 4151 at 2 keV to 2 MeV from HEAO 1," *Astrophysical Journal*, **279**, 555-562, 1984.
- Band, David L. and Jonathan N. Grindlay, "The Synchrotron-Self-Compton Process in Spherical Geometries: II Applications to Active Galactic Nuclei," *Astrophysical Journal*, **308**, 571-591, 1986.
- Barr, P. and R. F. Mushotzky, "Limits of x-ray variability in active galactic nuclei," *Nature*, **320**, 421-423, 1986.
- Barrett, H. H. and W. Swindell, *Radiological Imaging, Vol. 1 and 2*, Academic Press Inc., Orlando, Florida, 1981.

- Bassani, L., "The gamma-ray emission from the Seyfert galaxy NGC4151," *Astrophysics and Space Science*, **79**, 461-481, 1981.
- Bassani, L. and A. J. Dean, "Extragalactic Gamma-Ray Astronomy," *Space Science Reviews*, **35**, 367-398, 1983.
- Bassani, L. and A. J. Dean, "The Saturation of the Cosmic Diffuse Gamma-Ray Background by the Seyfert Galaxy Component," *Astrophysics and Space Science*, **100**, 457-460, 1984.
- Bassani, L. and A. J. Dean, "Gamma-ray observations and relativistic jets in active galactic nuclei," *Astronomy and Astrophysics*, **161**, 85-88, 1986.
- Bealle, J. H., W. K. Rose, W. Graf, K. M. Price, W. A. Dent, R. W. Hobbs, E. K. Conklin, B. L. Ulich, B. R. Dennis, C. J. Crannell, J. F. Dolan, K. J. Frost, and L. E. Orwig, "Radio and X-Ray Variability of the Nucleus of Centaurus A (NGC 5128)," *Astrophysical Journal*, **219**, 836-844, 1978.
- Begelman, Mitchell C., Roger D. Blandford, and Marin J. Rees, "Theory of extragalactic radio sources," *Reviews of Modern Physics*, **56**, 255-351, 1984.
- Begelman, Mitchell C., "Toward a Unified Theory of Active Galactic Nuclei," *Annals of the New York Academy of Sciences*, **470**, 51-70, 1986.
- Bignami, G. F., C. E. Fichtel, R. C. Hartman, and D. J. Thompson, "Galaxies and Gamma-Ray Astronomy," *Astrophysical Journal*, **232**, 649-658, 1979.
- Blandford, R. D. and R. L. Znajek, "Electromagnetic extraction of energy from Kerr black holes," *Monthly Notices of the Royal Astronomical Society*, **179**, 433-456, 1977.
- Blandford, R. D., "Theoretical Models of Active Galactic Nuclei," in *Active Galactic Nuclei*, ed. J. E. Dyson, pp. 281-299, Manchester University Press, Manchester, England, 1985.

- Caroli, E., R. C. Butler, G. DiCocco, P. P. Maggioli, L. Natalucci, and A. Spizzichino, "Coded Masks in X- and  $\gamma$ -ray Astronomy: the Problem of the Signal-to-Noise Ratio Evaluation," *Il Nuovo Cimento*, **7C**, 786-804, 1984.
- Cavaliere, Alfonso, Emanuele Giallongo, and Fausto Vagnetti, "From Local Active Galactic Nuclei to Early Quasars," *Astrophysical Journal*, **296**, 402-415, 1985.
- Coe, M. J., L. Bassani, A. R. Engel, and J. J. Quenby, "The hard X-ray spectrum of NGC4151," *Monthly Notices of the Royal Astronomical Society*, **195**, 214-244, 1981.
- Condon, J. J., "Confusion and Flux-Density Error Distributions," *Astrophysical Journal*, **188**, 279-286, 1974.
- Condon, J. J., "Statistical Counts of Faint Sources at 2380 MHz," *Astrophysical Journal*, **222**, 745-751, 1978.
- Cook, W. R., M. Finger, T. A. Prince, and E. C. Stone, "Gamma-Ray Imaging with a Rotating Hexagonal Uniformly Redundant Array," *IEEE Tran. Nuclear Science*, **NS-31**, 771-775, 1984.
- Cook, W. R., M. Finger, and T. A. Prince, "A Thick Anger Camera for Gamma-Ray Astronomy," *IEEE Transactions on Nuclear Science*, **NS-33**, 129-133, 1985.
- Cutri, Roc M. and Christopher W. McAlary, "A Statistical Study of the Relationship Between Galaxy Interactions and Nuclear Activity," *Astrophysical Journal*, **296**, 90-105, 1985.
- Danese, L., G. DeZotti, G. Fasano, and A. Franceschini, "Cosmological evolution of AGNs in the X-ray band," *Astronomy and Astrophysics*, **161**, 1-6, 1986.
- Dean, A. J. and D. Ramsden, "Extragalactic Gamma-Rays," *Philisophical Transactions of the Royal Society of London A*, **301**, 577-602, 1981.

- Dicke, R. H., "Scatter Hole Cameras for X-rays and Gamma-Rays," *Astrophysical Journal Letters*, **153**, L101-L106, 1968.
- Elvis, Martin, Andrzej Soltan, and William C. Keel, "Very Low Luminosity Active Galaxies and the X-Ray Background," *Astrophysical Journal*, **283**, 479-485, 1984.
- Fabbiano, G., L. Miller, G. Trinchieri, M. Longair, and M. Elvis, "An X-Ray Survey of a Complete Sample of 3CR Radio Galaxies," *Astrophysical Journal*, **277**, 115-131, 1984.
- Fabian, A. C., "X-Rays from Active Galactic Nuclei," *Physica Scripta*, **T7**, 129-133, 1984.
- Fabian, A. C., "X-Rays from Active Galactic Nuclei," in *Active Galactic Nuclei*, ed. J. E. Dyson, pp. 221-232, Manchester University Press, 1985.
- Fenimore, E. E. and T. M. Cannon, "Coded Aperture Imaging with Uniformly Redundant Arrays," *Applied Optics*, **17**, 337-347, 1978.
- Fenimore, E. E., "Coded Aperture Imaging: Predicted Performance and Uniformly Redundant Arrays," *Applied Optics*, **17**, 3562-3570, 1978.
- Fenimore, E. E., "Coded Aperture Imaging: the Modulation Transfer Function for Uniformly Redundant Arrays," *Applied Optics*, **19**, 2465-2471, 1980.
- Fichtel, C. E., G. A. Simpson, and D. J. Thompson, "Diffuse Gamma Radiation," *Astrophysical Journal*, **222**, 833-849, 1978.
- Finger, M. H. and T. A. Prince, "Hexagonal Uniformly Redundant Arrays for Coded-Aperture Imaging," in *Proceedings of the 19th International Cosmic Ray Conference*, vol. 3, pp. 295-298, 1985.
- Fukada, Y., S. Hayakawa, I. Kasahara, F. Makino, and Y. Tanaka, "Energy spectrum of diffuse component of cosmic soft gamma rays," *Nature*, **254**, 398-399, 1975.

- Gehrels, N., T. L. Cline, B. J. Teegarden, W. S. Paciesas, J. Tueller, P. Durouchoux, and J. M. Hameiry, "The gamma-ray spectrum of Centaurus A: a high-resolution observation between 70 keV and 8 MeV," *Astrophysical Journal*, **278**, 112-118, 1984.
- Giacconi, R., G. Branduardi, U. Briel, A. Epstein, D. Frabrant, E. Feigelson, W. Forman, P. Gorenstein, J. Grindlay, H. Gursky, F. R. Harnden Jr., J. P. Henry, C. Jones, E. Kellogg, D. Koch, S. Murray, E. Schreier, F. Seward, H. Tananbaum, K. Topka, L. Van Speybroeck, S. S. Holt, R. H. Becker, E. A. Boldt, P. J. Serlemitsos, G. Clark, C. Canizarer, T. Markert, R. Novick, D. Helfand, and K. Long, "The Einstein (HEAO 2) X-Ray Observatory," *Astrophysical Journal*, **230**, 540-550, 1979a.
- Giacconi, R., J. Bechtold, G. Branduardi, W. Forman, J. P. Henry, C. Jones, E. Kellogg, H. van der Lann, W. Liller, H. Marshall, S. S. Murray, J. Pye, E. Schreier, W. L. Sargent, F. Seward, and H. Tananbaum, "A High-Sensitivity X-Ray Survey Using the Einstein Observatory and the Discrete Source Contribution to the Extragalactic X-Ray Background," *Astrophysical Journal Letters*, **234**, L1-L7, 1979b.
- Gioia, Isabella M., Tommaso Maccacaro, Ruby E. Schild, John T. Stocke, James W. Liebert, Ivan J. Danziger, Daniel Kunth, and Jan Lub, "The Medium Sensitivity Survey: A New Sample of X-Ray Sources with Optical Identifications and the Revised Extragalactic Log N-Log S," *Astrophysical Journal*, **283**, 495-511, 1984.
- Gould, R. J., "Origin of Cosmic X Rays," *American Journal of Physics*, **35**, 376-393, 1967.
- Guilbert, P. W., A. C. Fabian, and M. J. Rees, "Spectral and variability constraints on compact sources," *Monthly Notices of the Royal Astronomical Society*, **205**, 593-603, 1983.
- Gunson, J. and B. Polychronopolos, "Optimum Design of a Coded Mask X-ray Telescope for Rocket Applications," *Mon. Not. R. Astron. Soc.*, **177**, 485-497, 1976.



- Hall, R. D., C. A. Meegan, G. D. Walraven, F. T. Djuth, and R. C. Haymes, "Detection of Nuclear Gamma Rays from Centaurus A," *Astrophysical Journal*, **210**, 631-641, 1976.
- Hazard, C. and S. Mitton, *Active galactic nuclei*, Cambridge University Press, Cambridge, 1977.
- Heffernan, Daniel M. and Richard L. Liboff, "A Model for the Diffuse Gamma-Ray Background," *Physics Letters*, **101A**, 305-307, 1984.
- Horstman, H. M., G. Cavallo, and E. Moretti-Horstman, "The X and Gamma Diffuse Background," *Rivista Del Nuovo Ciento*, **5**, 255-311, 1975.
- Kinzer, R. L., W. N. Johnson, and J. D. Kurfess, "A Balloon Observation of the Diffuse Cosmic X-Radiation Above 20 keV," *Astrophysical Journal*, **222**, 370-378, 1978.
- Kurfess, J. D., W. N. Johnson, R. L. Kinzer, G. H. Share, M. S. Strickman, M. P. Ulmer, D. D. Clayton, and C. S. Dyer, "The Oriented Scintillation Spectrometer Experiment for the Gamma-Ray Observatory," *Advances in Space Research*, **3**, 109-112, 1983.
- Lawrence, A., "Repeated X-ray flaring of NGC 4151," *Monthly Notices of the Royal Astronomical Society*, **192**, 83-94, 1980.
- Lawrence, Andrew and Martin Elvis, "Obscuration and the Various Kinds of Seyfert Galaxies," *Astrophysical Journal*, **256**, 410-426, 1982.
- Levine, A. M. and et al., "The HEAO 1 A-4 Catalog of High-Energy X-Ray Sources," *Astrophysical Journal Sup.*, **54**, 581-617, 1984.
- Lightman, A. P., R. Giacconi, and H. Tananbaum, "X-Ray Flares in NGC4151: A Thermal Model and Constraints on a Central Black Hole," *Astrophysical Journal*, **224**, 375-380, 1978.
- Maccacaro, T. and I. M. Gioia, "X-ray Sources of Cosmological Relevance," in *Early Evolution of the Universe and Its Present Structure*, ed. G. O. Abell and G. Chincarini, pp. 7-18, 1983.

- Maccacaro, Tommaso, Isabella M. Gioia, and John T. Stocke, "X-Ray Selected Quasars and Seyfert Galaxies: Cosmological Evolution Luminosity Function, and Contribution to the X-Ray Background," *Astrophysical Journal*, **283**, 486-494, 1984.
- Malkan, Matthew A., Bruce Margon, and Gary A. Chanan, "The Underlying Galaxies of X-Ray Selected Quasars," *Astrophysical Journal*, **280**, 66-78, 1984.
- Marshall, F. E., E. A. Boldt, S. S. Holt, R. B. Miller, R. F. Mushotzky, L. A. Rose, R. E. Rothschild, and P. J. Serlemitsos, "The Diffuse X-Ray Background Spectrum from 3 to 50 keV," *Astrophysical Journal*, **235**, 4-10, 1980.
- McConnel, M. L., D. J. Forrest, E. L. Chupp, and P. P. Dunphy, "A Coded Aperture Gamma Ray Telescope," *IEEE Trans. Nuclear Science*, **NS-29**, 155-159, 1982.
- Megan, C. A. and R. C. Haymes, "Evidence for Variability of Hard X-Rays From NGC4151," *Astrophysical Journal*, **233**, 510-513, 1979.
- Misner, Charles W., Kip S. Thorne, and John Archibald Wheeler, *Gravitation*, W. H. Freedman and Company, San Francisco, 1973.
- Mushotzky, R. F., F. E. Marshall, E. A. Bolt, S. S. Holt, and P. J. Serlemitsos, "HEAO 1 spectra of x-ray emitting Seyfert 1 galaxies," *Astrophysical Journal*, **235**, 377-385, 1980.
- Mushotzky, R. F. and F. E. Marshall, "X-Ray Observations of the Seyfert Galaxies Mkn 120 and MGC8-11-11," *Astrophysical Journal*, **239**, L5-L9, 1980.
- Mushotzky, R. F., "The X-ray Spectrum and Time Variability of Narrow Emission Line Galaxies," *Astrophysical Journal*, **256**, 92-102, 1982.
- Olive, Keith A. and Joseph Silk, "Diffuse Cosmic Gamma-Ray Background as a Probe of Cosmological Gravitino Regeneration and Decay," *Physical Review Letters*, **55**, 2362-2365, 1985.

- Perotti, F., A. Della Ventura, G. Sechi, G. Villa, G. Di Cocco, R. E. Baker, R. C. Butler, A. J. Dean, S. J. Martin, and D. Ramsden, "Low energy gamma-ray spectrum of NGC4151," *Nature*, **282**, 484-486, 1979.
- Perotti, F. et al., "Low-Energy Gamma-Ray Emission Close to CG 135+1," *Astrophysical Journal Letters*, **239**, L49-L52, 1980.
- Perotti, F., A. Della Ventura, G. Villa, G. Di Cocco, L. Bassani, R. C. Butler, J. N Carter, and A. J. Dean, "Detection of a Soft Gamma-Ray Emission from the Region of NGC4151," *Astrophysical Journal Letters*, **247**, L63-L66, 1981.
- Petre, R., R. F. Mushotzky, J. J. Krolik, and S. S. Holt, "Soft X-Ray Spectral Observations Of Quasars and High X-Ray Luminosity Seyfert Galaxies," *Astrophysical Journal*, **280**, 499-515, 1984.
- Piccinotti, G., R. F. Mushotzky, E. A. Holt, F. E. Marshall, P. J. Serlemitsos, and R. A. Shafer, "A Complete X-Ray Sample of the High-Latitude ( $|b| > 20$ ) Sky From HEAO 1 A-2: LogN-LogS and Luminosity Functions," *Astrophysical Journal*, **253**, 485-503, 1982.
- Pollock, A. M. and et al., "Search for Gamma-radiation from Extragalactic Objects Using a Likelihood Method," *Astronomy and Astrophysics*, **94**, 116-120, 1981.
- Reichert, G. A., R. F. Mushotzky, R. Petre, and S. S. Holt, "Soft X-Ray Spectral Observations of Low-Luminosity Active Galaxies," *Astrophysical Journal*, **296**, 69-89, 1985.
- Rees, Martin J., "Blackhole Models for Active Galactic Nuclei," *Ann. Rev. Astron. Astrophys.*, **22**, 471-506, 1984.
- Ricketts, M. J., B. A. Cooke, and K. A. Pounds, "X-ray transient source at high galactic latitude and suggested extragalactic identification," *Nature*, **259**, 536-547, 1976.

- Roos, Nico, "Galaxy Mergers and Active Nuclei I. The Luminosity Function," *Astrophysical Journal*, **294**, 479-485, 1985.
- Roos, Nico, "Galaxy Mergers and Active Nuclei II. Cosmological Evolution," *Astrophysical Journal*, **294**, 486-493, 1985.
- Rothschild, R., E. Boldt, S. Holt, P. Serlemitsos, G. Garmire, P. Agrawal, G. Rigler, S. Boyer, and M. Lampton, "The Cosmic X-Ray Experiment Aboard HEAO-1," *Space Science Instrumentation*, **4**, 269-301, 1979.
- Rothschild, R. E., R. F. Mushotzky, W. A. Baity, D. E. Gruber, J. L. Matteson, and L. E. Peterson, "2-165 keV Observations of Active Galaxies and the Diffuse Background," *Astrophysical Journal*, **269**, 423-437, 1983.
- Scheuer, P. A. G., "A Statistical Method for Analysing Observations of Faint Radio Stars," *Proceedings of the Cambridge Philosophical Society*, **53**, 764-773, 1957.
- Scheuer, P. A. G., "Fluctuations in the X-Ray Background," *Monthly Notices of the Royal Astronomical Society*, **166**, 329-337, 1974.
- Schmidt, Maarten and Richard F. Green, "Counts, Evolution, and Background Contribution of X-Ray Quasars and Other Extragalactic X-Ray Sources," *Astrophysical Journal*, **305**, 68-82, 1986.
- Schonfelder, V., U. Graser, and J. Daugherty, "Diffuse Cosmic and Atmospheric Gamma Radiation from Balloon Observations," *Astrophysical Journal*, **217**, 306-319, 1977.
- Schonfelder, V., F. Graml, and F. P. Penningsfeld, "The Vertical Component of 1-20 MeV Gamma Rays at Balloon Altitudes," *Astrophysical Journal*, **240**, 350-362, 1980.
- Schonfelder, V., "Discrete Gamma Ray Sources," in *Non-Solar Gamma-rays*, ed. R. Cowsik and R. D. Wills, Pergamon Press, Oxford, 1980.

- Setti, G., "Active Galactic Nuclei and Their Cosmological Evolution," in *MPE Report 184, X-ray and UV Emission from AGN*, ed. Joachim Trumper, pp. 243-253, 1984.
- Shafer, Richard Arrick, "Spatial Fluctuations in the Diffuse Cosmic X-Ray Background," in *Ph. D. Thesis, University of Maryland*, 1983.
- Skinner, G. K., "Imaging with Coded-Aperture Masks," *Nuclear Instruments and Methods*, **221**, 33-40, 1984.
- Stockton, Alan, "The Environment of QSOs," *Astrophysics and Space Science*, **118**, 487-497, 1986.
- Swanenburg, B. N., K. Bennett, G. F. Bignami, P. Caraveo, W. Hermsen, G. Kanbach, J. L. Mousnou, H. A. Mayer-Hasewander, J. A. Paul, B. Sacco, L. Scarsi, and R. D. Wills, "COS B observations of high-energy gamma radiation from 3C273," *Nature*, **275**, p. 298, 1978.
- Tennant, Allyn F. and R. F. Mushotzky, "The Absence of Rapid Variability in Active Galaxies," *Astrophysical Journal*, **264**, 92-104, 1983.
- Trombka, J. I., C. S. Dyer, L. G. Evans, M. J. Bielefeld, S. M. Seltzer, and A. E. Metzger, "Reanalysis of the Apollo Cosmic Gamma-Ray Spectrum in the 0.3 to 10 MeV Energy Region," *Astrophysical Journal*, **212**, 925-935, 1977.
- Ulrich, M. H., "Line Variability in Active Nuclei and the Structure of the Broad-Line Region," *Annals of the New York Academy of Sciences*, **422**, 291-302, 1984.
- Urry, Claudia Megan, "X-Ray and Ultraviolet Observations of BL Lacertae Objects," in *Ph.D. Thesis, John Hopkins University*, Baltimore, Maryland, 1984.
- Wandel, Amri and Amos Yahil, "Universal Mass-Luminosity Relation for Quasars and Active Galactic Nuclei?," *Astrophysical Journal Letters*, **295**, L1-L4, 1985.

- Webber, W. R., J. A. Lockwood, and G. A. Simpson, "A New Look at the Diffuse Gamma-Ray Spectrum Between 0.1 and 200 MeV," in *Proceeding of the 17th International Cosmic-Ray Conference*, vol. 1, pp. 247-250, 1981.
- White, R. Stephen, Bruce Dayton, Sin M. Moon, Janes M. Ryan, Robert B. Wilson, and Allen D. Zych, "Cosmic Diffuse Gamma Rays From 2 to 25 MeV," *Astrophysical Journal*, **218**, 920-927, 1977.
- White, R. S., B. Dayton, R. Gibbons, J. L. Long, E. M. Zanrosso, and A. D. Zych, "Observation and search for gamma rays 1-20 MeV from the Crab, NGC4151, Cygnus X-3, CG135+1 and 3C273," *Nature*, **284**, 608-610, 1980.
- Zdziarski, Andrzej, "Effect of Compton Scattering on the Pair Annihilation and Bremsstrahlung Spectra," *Physica Scripta*, **T7**, 124-126, 1984.
- Zdziarski, Andrzej A. and Alan P. Lightman, "NonThermal Electron-Positron Pair Production and the "Universal" X-ray Spectrum of Active Galactic Nuclei," *Astrophysical Journal Letters*, **294**, L79-L83, 1985.